



Implementation K-Means Algorithm Determine the Recovery Rate Of Covid-19 Patients In Indonesia

Mentari Tri Indah Rahmayani¹, Nani Hidayati²

¹ Sekolah Tinggi Agama Islam Negeri Bengkalis, Indonesia

² STIKOM Tunas Bangsa Pematangsiantar, Indonesia

E-mail : mentari.tri@gmail.com¹, nanihidayati3@gmail.com²

ARTICLE INFO

ABSTRACT

Article history:

Received: Mar 08, 2022

Revised: Apr 22, 2022

Accepted: May 13, 2022

Keywords:

COVID-19,
Clustering,
Data Mining

In Indonesia, the number of COVID-19 sufferers continues to increase every time, data from the Indonesian National Disaster Management (BPBN) shows the number shows the number of deaths as many as 100,636 people and the recovery rate as many as 2,907,920 people. In this study, the authors applied the k-means clustering algorithm to group the areas with the lowest cure rate (C1), moderate cure rate (C2) and high cure rate (C3). The purpose of this study is to group areas in Indonesia with the highest cure rates and to find out the similarities in data management using manual calculations and using Rapid Miner Software. The results show that from 34 provinces in Indonesia 30 provinces are in the low cluster, namely, Aceh, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, Bangka Belitung, Riau Islands, DI Jogjakarta, Banten, Bali, NTB, NTT, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, Central Sulawesi, South Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, North Maluku, Papua and West Papua, with a moderate cure rate cluster of 2 provinces, namely Central Java and East Java, and the cluster with the highest cure rate was 2 provinces, namely DKI Jakarta and West Java.

Copyright © 2022 Jurnal Mantik.
All rights reserved.

1. Introduction

Throughout world history, 2020 was the worst year for countries in the world. The world is currently experiencing a pandemic, namely the spread of the Coronavirus Disease of 2019 (COVID-19). COVID-19 is a virus that is very easily transmitted from children to the elderly who can be exposed to the COVID-19 virus [1]. In addition, COVID-19 can take lives.

Indonesia is one of the many countries that have been exposed to COVID-19, it is recorded that Indonesia is in 14th place with a total of 4,205,705 cases. This makes the Indonesian people restless because they have to think about health and also the declining economy. From statistical data from the Indonesian National Disaster Management Agency (BPBN) on August 3, 2021, it shows the number of deaths is 100,636 people and the recovery rate is 2,907,920 people.

The important role of data mining in preventing the spread of this virus. Along with the distribution of the roles of health care epidemiologists, soft-copy (electronic) health data are also evolving [2]. Data mining can also be interpreted as extracting new information taken from free chunks of data that helps in decision making [3].

In this study, the author uses the K-Means Clustering method. This algorithm was chosen because it has been tested and is often used by researchers for grouping [4].

K-Means Clustering is a grouping of records, observations, or attention and forms a class of objects that have similarities [5]. In addition, clustering is the process of grouping data sets into several groups so that objects in one group have many similarities and have many differences with objects in other groups. The



differences and similarities are usually based on the attribute values of the object and can also be in the form of distance calculations [6].

In this study, the authors used data on COVID-19 patients in Indonesia by grouping each province in Indonesia into 3 clusters, namely provinces with low cure rates (C1), moderate cure rates (C2) and high cure rates (C3). The purpose of this study is to group provinces in Indonesia into certain clusters, namely provinces with low, medium and high cure rates so that data can be obtained that can assist the government in determining future policies.

2. Method

The methods in this research are as follows:

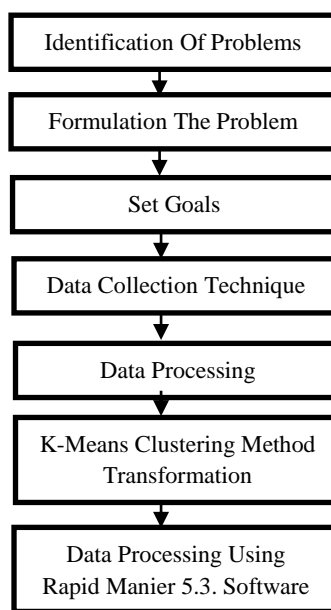


Figure 1. Research Flowchart

a. Problem identification

In determining the formulation of the problem, the author refers to the number of people who are exposed to COVID-19, especially the State of Indonesia and how to use the data that is already available to obtain new findings that can later be used by certain parties.

b. Problem Formulation

Formulation of the problem in this research is

- 1). How to determine the cure rate of COVID-19 sufferers in provinces in Indonesia based on low, medium and high categories?
- 2). How to get new knowledge related to the grouping of healing rates of COVID-19 sufferers in Indonesia?

c. Setting Goals

The objectives of this research are to:

- 1). Generate Clusters from data on recovering COVID-19 sufferers, especially provinces in Indonesia.
- 2). Knowing the provinces with the lowest, medium and high cure rates.
- 3). Comparing the results of Clustering with calculations using the K-Means Algorithm and Using Rapid Miner 5.3 software.

d. Data Collection Techniques

The data collection step is done by looking at the data. From statistical data from the Indonesian National Disaster Management Agency (BPBN) on August 3, 2021 and with literature studies taken

from various sources, namely articles, scientific journals about the K-Means Clustering method, as well as other supporting reading materials [7].

e. Data Processing

In making it easier for the author of manual data processing techniques assisted by Microsoft Excel.

f. Data processing using the K-Means Clustering Method

The data processing technique uses the K-means technique which is a grouping technique that works based on Partitioned Clustering [8]. The first stage to perform an algorithm comparison is to determine the data object to be processed, the second stage is to perform automatic data separation, training and testing through data split operators. The third stage is the data mining extraction process for the data sets that have been prepared previously [9]

g. Data Processing Using Rapid Manier 5.3. Software

The last stage is testing the results of data processing using RapidMiner Software to find out the equations for calculating the manual method and using Rapid Miner software [10].

3. Result and Discussion

In the process of calculating the data manually, the data used is data on healing patients with COVID-19 from March 2020 - August 3, 2021, as many as 34 provinces in Indonesia. The following data on Healing in Indonesia (Source: BNPB Indonesia).

TABEL 1
HEALING DATA FROM PATIENTS COVID-19

No	Provinsi	Total Healing
1	Aceh	16253
2	Sumatera Utara	35752
3	Sumatera Barat	52132
4	Riau	73327
5	Jambi	13017
6	Sumatera Selatan	29938
7	Bengkulu	10869
8	Lampung	21025
9	Bangka Belitung	22941
10	Kepulauan Riau	28815
11	Dki Jakarta	652207
12	Jawa Barat	398488
13	Jawa Tengah	253479
14	Di Jogjakarta	64913
15	Jawa Timur	184213
16	Banten	56260
17	Bali	53448
18	Nusa Tenggara Barat	13747
19	Nusa Tenggara Timur	18988
20	Kalimantan Barat	16417
21	Kalimantan Tengah	22565
22	Kalimantan Selatan	35643
23	Kalimantan Timur	79814
24	Kalimantan Utara	13329
25	Sulawesi Utara	16256
26	Sulawesi Tengah	13739
27	Sulawesi Selatan	65470
28	Sulawesi Tenggara	11390
29	Gorontalo	59254
30	Sulawesi Barat	5948
31	Maluku	8095
32	Maluku Utara	5419
33	Papua	11867



No	Provinsi	Total Healing
34	Papua Barat	11988

Algorithm Manual Calculation K-means Clustering

- Determining the amount of data that will be in the cluster, in this study the sample data for healing COVID-19 patients used were 34 provinces in Indonesia.
- Determining the value of K, the number of clusters used is 3 clusters, namely low, medium and high clusters.
- Determine the initial Centroid value randomly, low cluster (C1), medium cluster (C2) and high cluster (C3).

Here is the initial centroid value at iteration 1

TABEL 2
INITIAL CENTROID OF ITERATION 1

CI	Minimum	5419
C2	Average	18988
C3	Maximum	652207

- After the initial centroid is determined, then calculate the distance of each data from the center of the cluster to the center point in the first cluster. Here is the calculation: $D(1.1) = \sqrt{(5419 - 16253)^2} = 10,834$

The calculation is done until D (1.34), then the calculation in cluster 2 with the center point of the centroid:

$D(2.1) = \sqrt{(18988 - 16253)^2} = 2735$ The calculation is done until D (2.34), next calculation on cluster 3 with center point centroid :

$D(3.1) = \sqrt{(652207 - 16253)^2} = 635954$ The calculation is done until D (3.34).

Calculation of the centroid value continues. So that the results obtained from the calculation of data with the center point of the centroid of each cluster are as follows:

TABEL 3
CALCULATION OF THE CENTROID IN THE CLUSTER (ITERATION I)

No	Provinsi	Total Healing	C1	C2	C3
1	Aceh	16253	10834	2735	635954
2	Sumatera Utara	35752	30333	16764	616455
3	Sumatera Barat	52132	46713	33144	600075
4	Riau	73327	67908	54339	578880
5	Jambi	13017	7598	5971	639190
6	Sumatera Selatan	29938	24519	10950	622269
7	Bengkulu	10869	5450	8119	641338
8	Lampung	21025	15606	2037	631182
9	Bangka Belitung	22941	17522	3953	629266
10	Kepulauan Riau	28815	23396	9827	623392
11	Dki Jakarta	652207	646788	633219	0
12	Jawa Barat	398488	393069	379500	253719
13	Jawa Tengah	253479	248060	234491	398728
14	Di Jogjakarta	64913	59494	45925	587294
15	Jawa Timur	184213	178794	165225	467994



No	Provinsi	Total Healing	C1	C2	C3
16	Banten	56260	50841	37272	595947
17	Bali	53448	48029	34460	598759
18	Nusa Tenggara Barat	13747	8328	5241	638460
19	Nusa Tenggara Timur	18988	13569	0	633219
20	Kalimantan Barat	16417	10998	2571	635790
21	Kalimantan Tengah	22565	17146	3577	629642
22	Kalimantan Selatan	35643	30224	16655	616564
23	Kalimantan Timur	79814	74395	60826	572393
24	Kalimantan Utara	13329	7910	5659	638878
25	Sulawesi Utara	16256	10837	2732	635951
26	Sulawesi Tengah	13739	8320	5249	638468
27	Sulawesi Selatan	65470	60051	46482	586737
28	Sulawesi Tenggara	11390	5971	7598	640817
29	Gorontalo	59254	53835	40266	592953
30	Sulawesi Barat	5948	529	13040	646259
31	Maluku	8095	2676	10893	644112
32	Maluku Utara	5419	0	13569	646788
33	Papua	11867	6448	7121	640340
34	Papua Barat	11988	6569	7000	640219

Then calculate the shortest distance using the Eulidean Distance. The following are the results of the calculation of the shortest distance from this study.

TABEL 4
SHORTEST DISTANCE VALUE

Shortest Distance	C1	C2	C3
2735	C2	1	
16764	C2	1	
33144	C2	1	
54339	C2	1	
5971	C2	1	
10950	C2	1	
5450	C1	1	
2037	C2	1	
3953	C2	1	
9827	C2	1	
0	C3		1
253719	C3		1
234491	C2	1	
45925	C2	1	
165225	C2	1	
37272	C2	1	
34460	C2	1	
5241	C2	1	



Shortest Distance	C1	C2	C3
0	C2	1	
2571	C2	1	
3577	C2	1	
16655	C2	1	
60826	C2	1	
5659	C2	1	
2732	C2	1	
5249	C2	1	
46482	C2	1	
5971	C1	1	
40266	C2	1	
529	C1	1	
2676	C1	1	
0	C1	1	
6448	C1	1	
6569	C1	1	

So that the calculation results from the cluster center in iteration I are:

TABLE 5
CLUSTERING RESULTS

Cluster	Results
C1	7
C2	25
C3	2

- e. The next step is back in steps 4-5 until the centroid result from iteration I with the next centroid value is the same and does not change in the position of the cluster.
- f. To calculate the latest center point using the results of the cluster in iteration I. The following is the calculation of the new center point of the cluster in iteration II:

$$C1 = \frac{5450+5971+529+2676+0+6448+6569}{7}$$

$$= 3949$$

$$C2 = \frac{2735+16764+33144+54339+5971+10950+2037+3953+9827+234491+45925+165225+37272+34460+5241+0+2571+3577+16655+60826+5659+2732+5249+46482+40266}{25}$$

$$C3 = \frac{0+253719}{2}$$

TABLE 6
RESULTS OF GROUPING THE HEALING RATE OF COVID-19 PATIENTS

No	Provinsi	Total Healing	Results	C1	C2	C3
1	Aceh	16253	C1	1		
2	Sumatera Utara	35752	C1	1		
3	Sumatera Barat	52132	C1	1		
4	Riau	73327	C1	1		
5	Jambi	13017	C1	1		
6	Sumatera Selatan	29938	C1	1		
7	Bengkulu	10869	C1	1		
8	Lampung	21025	C1	1		
9	Bangka Belitung	22941	C1	1		
10	Kepulauan Riau	28815	C1	1		
11	Dki Jakarta	652207	C3			1
12	Jawa Barat	398488	C3			1
13	Jawa Tengah	253479	C2		1	



No	Provinsi	Total Healing	Results	C1	C2	C3
14	Di Jogjakarta	64913	C1	1		
15	Jawa Timur	184213	C2		1	
16	Banten	56260	C1	1		
17	Bali	53448	C1	1		
18	Nusa Tenggara Barat	13747	C1	1		
19	Nusa Tenggara Timur	18988	C1	1		
20	Kalimantan Barat	16417	C1	1		
21	Kalimantan Tengah	22565	C1	1		
22	Kalimantan Selatan	35643	C1	1		
23	Kalimantan Timur	79814	C1	1		
24	Kalimantan Utara	13329	C1	1		
25	Sulawesi Utara	16256	C1	1		
26	Sulawesi Tengah	13739	C1	1		
27	Sulawesi Selatan	65470	C1	1		
28	Sulawesi Tenggara	11390	C1	1		
29	Gorontalo	59254	C1	1		
30	Sulawesi Barat	5948	C1	1		
31	Maluku	8095	C1	1		
32	Maluku Utara	5419	C1	1		
33	Papua	11867	C1	1		
34	Papua Barat	11988	C1	1		

Grouping using Rapid Miner 5.3 software can be seen from the following figure: Cluster results from 34 provinces in Indonesia were obtained for cluster 0 which is a low cure rate, there are 30 provinces, cluster 1, which is a moderate cure rate, there are 2 provinces and cluster 2, which is a high cure rate, there are 2 provinces.

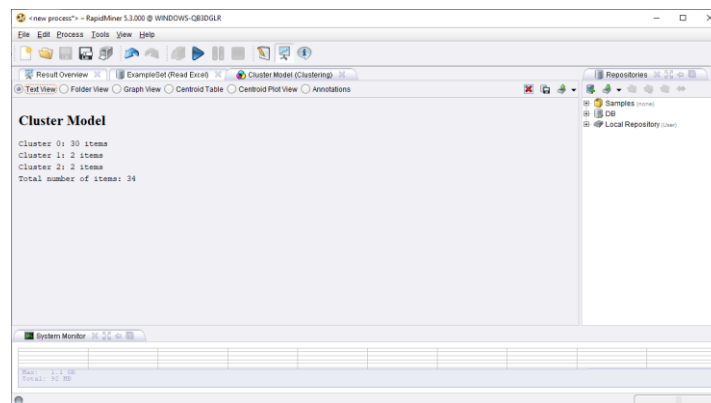


Figure 2 Display Cluster Model with Rapid Miner Software 5.3

The final result of the centroid value shows the results of cluster 0 which is 29620.633, cluster 1 is 218846 and cluster 2 is 525347.500

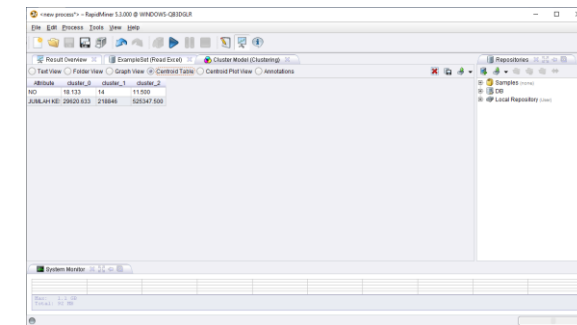


Figure 3 Centroid Value with Rapidminer Software 5.3

From the results of grouping using Rapid Miner 5.3 software, 30 provinces were obtained cluster 0, cluster 1 as many as 2 provinces and cluster 2 as many as 2 provinces.

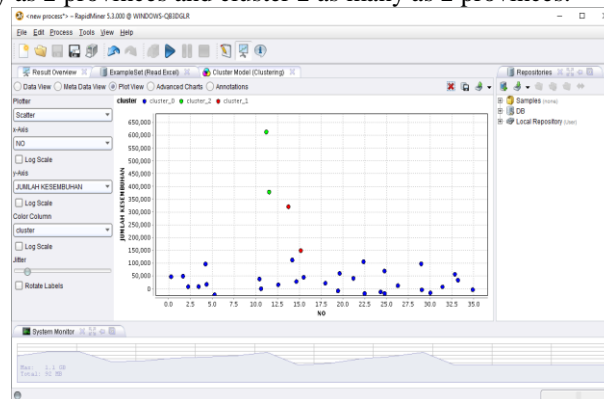


Figure 4 Grouping Results using Rapidminer Software

4. Conclusion

The conclusions obtained from the results of research on k-means clustering of the total recovery rate of COVID-19 sufferers, especially the provinces in Indonesia are: The k-means algorithm can help researchers to classify the cure rate of COVID-19 sufferers. From the results of clustering, 30 provinces were found in the low cluster, namely, Aceh, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, Bangka Belitung, Riau Islands, DI Jogjakarta, Banten, Bali, NTB, NTT, Kalimantan. West, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, Central Sulawesi, South Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, North Maluku, Papua and West Papua, with a moderate cure rate cluster of 2 provinces namely Central Java and East Java and the clusters with the highest cure rate were 2 provinces, namely DKI Jakarta and West Java. Manual calculations using the k-means algorithm and the use of Rapid Minar software produce the same value.

5. References

- [1] N. Mirantika, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat," NUANSA Inform., vol. 15, no. 2, pp. 92–98, Aug. 2021, doi: 10.25134/nuansa.v15i2.4321.
- [2] O. P. Barus and A. Tehja, "PREDIKSI KESEMBUHAN PASIEN COVID-19 DI INDONESIA MELALUI TERAPI MENGGUNAKAN METODE NAÏVE BAYES," Inf. Syst. Dev., vol. 6, no. 2, p. 8.
- [3] N. Chafid and I. A. Wibowo, "IMPLEMENTASI DATA MINING UNTUK CLUSTERING DAERAH PENYEBARAN PENYAKIT DEMAM BERDARAH DI KOTA TANGERANG SELATAN MENGGUNAKAN ALGORITMA K-MEANS," vol. 3, no. 1, p. 13, 2018.



- [4] Y. F. S. Y. Damanik, S. Sumarno, I. Gunawan, D. Hartama, and I. O. Kirana, "Penerapan Data Mining Untuk Pengelompokan Penyebaran Covid-19 Di Sumatera Utara Menggunakan Algoritma K-Means," *J. Ilmu Komput. Dan Inform.*, vol. 1, no. 2, Nov. 2021, doi: 10.54082/jiki.13.
- [5] N. Noviyanto, "Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia," *Paradig. - J. Komput. Dan Inform.*, vol. 22, no. 2, pp. 183–188, Sep. 2020, doi: 10.31294/p.v22i2.8808.
- [6] M. T. I. Rahmayani, "ANALISIS CLUSTERING TINGKAT KEPARAHAN PENYAKIT PASIEN MENGGUNAKAN ALGORITMA K-MEANS (STUDI KASUS DI PUSKESMAS BANDAR SEIKIJANG)," no. 2, p. 5, 2018.
- [7] Y. Aswan, S. Defit, and G. W. Nurcahyo, "Algoritma K-Means Clustering dalam Mengklasifikasi Data Daerah Rawan Tindak Kriminalitas (Polres Kepulauan Mentawai)," *J. Sistim Inf. Dan Teknol.*, pp. 245–250, Aug. 2021, doi: 10.37034/jsisfotek.v3i4.73.
- [8] R. Handoyo, "PERBANDINGAN METODE CLUSTERING MENGGUNAKAN METODE SINGLE LINKAGE DAN K - MEANS PADA PENGELOMPOKAN DOKUMEN," vol. 15, no. 2, p. 10, 2014.
- [9] I. D. Iskandar, M. W. Pertiwi, M. Kusmira, and I. Amirulloh, "KOMPARASI ALGORITMA CLUSTERING DATA MEDIA ONLINE PADA PROSES BISNIS," vol. 2, no. 3, p. 8, 2018.
- [10] W. J. Mawaddah, I. Gunawan, and I. P. Sari, "Implementasi Algoritma Data Mining untuk Klustering Data Hasil Panen Kelapa Sawit," vol. 1, no. 1, p. 12.

