Implementation of the K-Medoids Algorithm for Data Clustering of Covid 19 Cases in West Java

Ririn Restu Aria

Universitas Bina Sarana Informatika, Indonesia

ririn.rra@bsi.ac.id

Abstract

The Covid 19 pandemic has hit Indonesia for almost 15 months since March 2020. The virus has spread to all provinces in Indonesia. Various efforts were made to be able to reduce or prevent the spread of the coronavirus, including the implementation of the PSBB in various areas including in West Java province. In this study, the objective of this research is to cluster the data on cases of Covid 19 in West Java which are recapitulated daily based on districts/cities that occurred on May 20, 2021. For the clustering process, the K-medoids algorithm is used which determines 3 clusters based on the variables used, namely discarded close contact, suspects discarded, probable completed, probable died, totally positive, positive recovered, and positive died. For data processing, a calculation analysis was carried out using the stages in the K-medoids algorithm and the Rapidminer application with high cluster mapping of 6 districts/cities, medium clusters there were 19 districts/cities, while low clusters had 2 districts/cities. The results of the analysis are expected to provide information about the distribution and mapping of clusters in West Java province.

*Keywords: Covid-19,Clustering, K-medoids Algorithm*

1. Introduction

2019 in December, the coronavirus outbreak began to attack residents of Wuhan province [1] through the respiratory tract. The spread of the virus is fast enough to infect all countries in the world, including Indonesia, which occurred in March 2020 [2] [3]. The spread of covid 19 has a significant impact on various sectors of the economy, education, tourism, and other sectors [4] [5] [6] which requires people to carry out social distancing and restrictions on activities carried out by a large number of people. One province that has fairly high data on the spread of Covid 19 cases is West Java province with a total of 302,335 confirmed people, 28,938 people in isolation / in care, 269,351 people who have been isolated/recovered and 4,046 people died [7]. In the study, applying the K-Medoids algorithm to cluster the dataset used in the case of Covid 19, obtained based on districts/cities in West Java, which are grouped from the concept of data mining to find the partial clustering pattern used to find k-clusters of data collected. best characterizes the objects in that data set. Several previous studies that discussed Covid 19 used the K-medoids algorithm. Sindi et al conducted a study in 2020 by grouping the spread of covid 19 using the K-medoids algorithm in the province of Indonesia which is divided into 3 clusters [8]. Subsequent researchers Samudi et al who discussed the use of instructional media in the Covid 19 pandemic [9] and Windarto et al, who combined clustering and classification methods for Covid 19 cases in Indonesia [10].

2. Research Methodology
2.1. Data on the Distribution of Covid Cases 19

When conducting research, case data related to the distribution of cases that occur in West Java based on districts / cities are needed using the dataset from Discarded close contact, suspect discarded, probable completion, Probable Death, positive total, positive
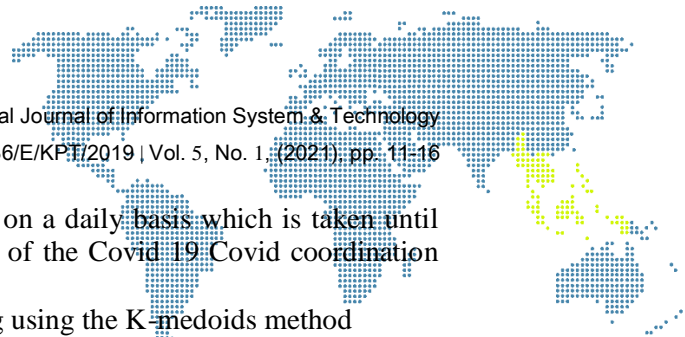
recovery and positive death which are recapitulated on a daily basis which is taken until 20 May 2021 which was obtained from the website of the Covid 19 Covid coordination and information center in West Java Province [7].

Several stages were carried out in the data processing using the K-medoids method

a) Data Selection

The data set from this study was obtained from the Pikobar website [7] by following the data requirements required for data processing.

b) Dataset Selection

Researchers determine the data be processed based on the date grouped by districts/cities in West Java, which is cumulative according to the date selected by the researcher.

2.2. Data Mining

Data mining is a method used to process information obtained in a database that can be used for certain purposes based on the algorithm that will be used [11] [12] [13].

2.3. Clustering

Is the first step that will be used for grouping data that is seen based on the similarity of objects [8]. When using clustering, the equations are used to determine the distance to the algorithm, which means that the initial data selection can be determined randomly [14].

2.4. K-Medoids Algorthm

It is an algorthm that is used to determine k which is determined as an object representative to minimize the number of object inequalities. In the K-medoids algorithm, we must first determine the center point of the cluster that will be used in data grouping. The steps used to complete the K-Medoids calculation

a) Determine the initial centroid that is randomly selected from the existing data set

b) Count non-medoids objects to the closest cluster based on the Euclidean distance.

c) Randomly assign objects in each cluster as candidate data objects for the new medoids

d) Calculate the distance of each object from each cluster

e) Perform the calculation of the Total Deviation (S) with the formula S = new total distance value - old total distance. If the value of S <0 then replace the object with cluster data to get k new objects as medoids

3. Results and Discussion

a) Selecting Covid 19 data in West Java based on 27 districts/cities categorized based on the parameters to be calculated, namely the number of confirmed positives, the number of cured, the number of people who died, the number of isolations in care, and the number in close contact. The results of normalization can be seen. in table 1.

Table 1. Covid Case Data 18 West Java

Name of Regency / City	Closely Discarded Contact	Discarded suspension	Probable Done	Probable Death	Total Positive	Total Heal	Positive died
Kab. Bogor	0	5409	118	291	17862	12876	172
Kab. Sukabumi	8924	7645	208	134	4893	4580	34
Kab. Cianjur	1845	2188	27	28	3593	2697	2
Kab. Bandung	26068	4299	0	29	13116	10144	43
Kab. Garut	13287	5031	0	0	10638	8331	143
Kab. Tasikmalaya	2188	1829	4	9	2679	2196	60
Kab. Ciamis	45998	42	3	0	4197	2915	55
Kab. Kuningan	18552	1303	22	9	5675	4336	16
Kab. Cirebon	8267	381	0	24	8741	6297	194
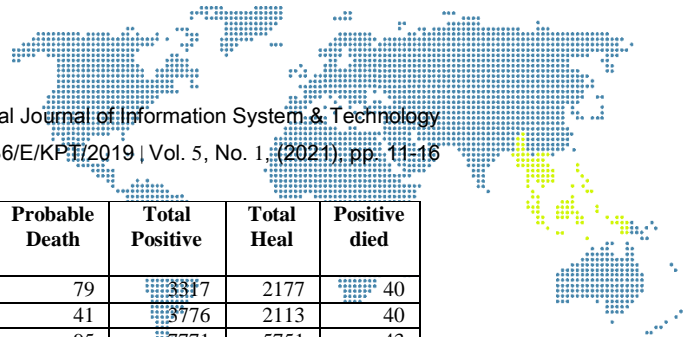
Name of Regency / City	Closely Discarded Contact	Discarded suspension	Probable Done	Probable Death	Total Positive	Total Heal	Positive died
Kab. Majalengka	5355	1444	0	79	3317	2177	40
Kab. Sumedang	2892	956	0	41	3776	2113	40
Kab. Indramayu	13361	4062	0	95	7771	5751	43
Kab. Subang	923	6102	35	0	5177	3661	14
Kab. Purwakarta	2825	1554	1	23	4546	3650	59
Kab. Karawang	17705	7461	0	87	19057	17200	237
Kab. Bekasi	22253	10928	675	67	25661	24897	60
Kab. Bandung Barat	1536	865	5	45	6885	4371	37
Kab. Pangandaran	1015	2545	1	4	2355	1024	2
Kota Bogor	6956	3611	11	89	15546	13346	120
Kota Sukabumi	8055	9711	2	25	4477	3398	42
Kota Bandung	19151	15784	150	24	18943	17473	89
Kota Cirebon	5273	2510	1	1	5354	4270	36
Kota Bekasi	38818	17313	1134	185	41908	39038	218
Kota Depok	45976	18206	0	7	48585	44665	322
Kota Cimahi	8685	2050	0	14	5141	4030	54
Kota Tasikmalaya	1131	1443	75	10	6089	4709	32
Kota Banjar	1946	656	19	9	1519	769	24

Source :https://pikobar.jabarprov.go.id

b) Normalized the covid 19 datasets in West Java based on districts/cities. The results of normalization can be seen in table 2.

Table 2. Normalization of data

Name of Regency / City	Closely Discarded Contact	Discarded suspension	Probable Done	Probable Death	Total Positive	Total Heal	Positive died
Kab. Bogor	0.00	0.30	0.10	1.00	0.35	0.28	0.53
Kab. Sukabumi	0.19	0.42	0.18	0.46	0.07	0.09	0.10
Kab. Cianjur	0.04	0.12	0.02	0.10	0.04	0.04	0.00
Kab. Bandung	0.57	0.23	0.00	0.10	0.25	0.21	0.13
Kab. Garut	0.29	0.27	0.00	0.00	0.19	0.17	0.44
Kab. Tasikmalaya	0.05	0.10	0.00	0.03	0.02	0.03	0.18
Kab. Ciamis	1.00	0.00	0.00	0.00	0.06	0.05	0.17
Kab. Kuningan	0.40	0.07	0.02	0.03	0.09	0.08	0.04
Kab. Cirebon	0.18	0.02	0.00	0.08	0.15	0.13	0.60
Kab. Majalengka	0.12	0.08	0.00	0.27	0.04	0.03	0.12
Kab. Sumedang	0.06	0.05	0.00	0.14	0.05	0.03	0.12
Kab. Indramayu	0.29	0.22	0.00	0.33	0.13	0.11	0.13
Kab. Subang	0.02	0.33	0.03	0.00	0.08	0.07	0.04
Kab. Purwakarta	0.06	0.08	0.00	0.08	0.06	0.07	0.18
Kab. Karawang	0.38	0.41	0.00	0.30	0.37	0.37	0.73
Kab. Bekasi	0.48	0.60	0.60	0.23	0.51	0.55	0.18
Kab. Bandung Barat	0.03	0.05	0.00	0.15	0.11	0.08	0.11
Kab. Pangandaran	0.02	0.14	0.00	0.01	0.02	0.01	0.00
Kota Bogor	0.15	0.20	0.01	0.31	0.30	0.29	0.37
Kota Sukabumi	0.18	0.53	0.00	0.09	0.06	0.06	0.13
Kota Bandung	0.42	0.87	0.13	0.08	0.37	0.38	0.27
Kota Cirebon	0.11	0.14	0.00	0.00	0.08	0.08	0.11
Kota Bekasi	0.84	0.95	1.00	0.64	0.86	0.87	0.68
Kota Depok	1.00	1.00	0.00	0.02	1.00	1.00	1.00
Kota Cimahi	0.19	0.11	0.00	0.05	0.08	0.07	0.16
Kota Tasikmalaya	0.02	0.08	0.07	0.03	0.10	0.09	0.09
Kota Banjar	0.04	0.03	0.02	0.03	0.00	0.00	0.07

c) Determine the initial Centroid value of Medoids and calculate the proximity of the distance between the clusters

Table 3. Medoids Initial Centroid Values

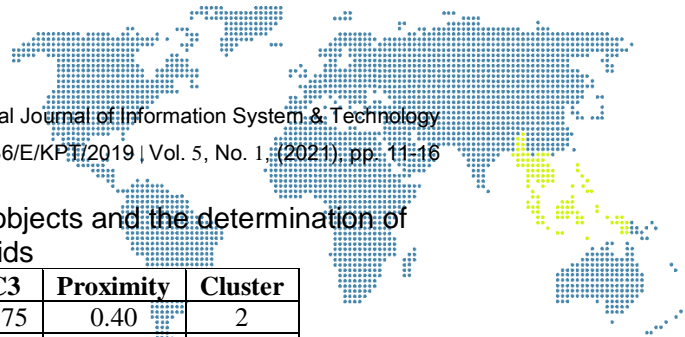| Kab. Bogor	0.00	0.00	0.17	1.00	0.35	0.02	0.53
Kab. Tasikmalaya	0.00	0.00	0.01	0.03	0.02	0.00	0.18
Kota Bekasi	0.04	0.02	0.00	0.64	0.86	0.05	0.68

13

Table 4. The value of the proximity between objects and the determination of clusters in Medoids

Name of Regency / City	C1	C2	C3	Proximity	Cluster
Kab. Bogor	0.40	1.16	0.75	0.40	2
Kab. Sukabumi	0.88	0.67	1.09	0.67	2
Kab. Cianjur	1.11	0.23	1.19	0.23	2
Kab. Bandung	1.19	0.69	1.14	0.69	1
Kab. Garut	1.12	0.53	1.02	0.53	2
Kab. Tasikmalaya	1.10	0.11	1.15	0.11	2
Kab. Ciamis	1.50	1.00	1.49	1.00	1
Kab. Kuningan	1.20	0.44	1.22	0.44	2
Kab. Cirebon	0.98	0.49	0.91	0.49	2
Kab. Majalengka	0.92	0.29	1.06	0.29	2
Kab. Sumedang	1.02	0.15	1.10	0.15	2
Kab. Indramayu	0.91	0.50	1.01	0.50	2
Kab. Subang	1.20	0.37	1.23	0.37	2
Kab. Purwakarta	1.05	0.14	1.09	0.14	2
Kab. Karawang	1.00	0.97	0.85	0.85	1
Kab. Bekasi	1.34	1.23	1.29	1.23	1
Kab. Bandung Barat	0.99	0.19	1.05	0.19	2
Kab. Pangandaran	1.19	0.23	1.25	0.23	2
Kota Bogor	0.82	0.57	0.79	0.57	2
Kota Sukabumi	1.19	0.57	1.23	0.57	2
Kota Bandung	1.40	1.10	1.30	1.10	1
Kota Cirebon	1.14	0.22	1.16	0.22	2
Kota Bekasi	1.85	2.16	1.79	1.79	3
Kota Depok	2.14	2.15	1.81	1.81	3
Kota Cimahi	1.09	0.24	1.12	0.24	2
Kota Tasikmalaya	1.10	0.18	1.14	0.18	2
Kota Banjar	1.14	0.13	1.21	0.13	2

Besides using manual calculations, the K-Medoids algorithm can also use calculations using the Rapidminer application. Based on the research dataset processed using Rapidminer, it is found that the number of clusters consists of 3 clusters, namely cluster 0 = 6 districts / cities, cluster 1 = 19 districts / cities and cluster 2 = 2 districts / cities.
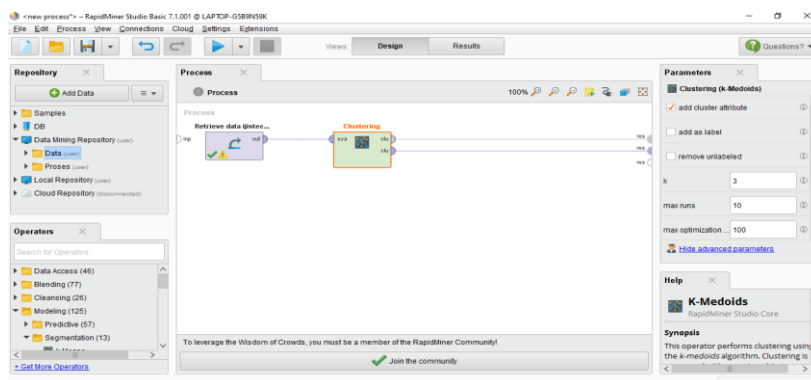

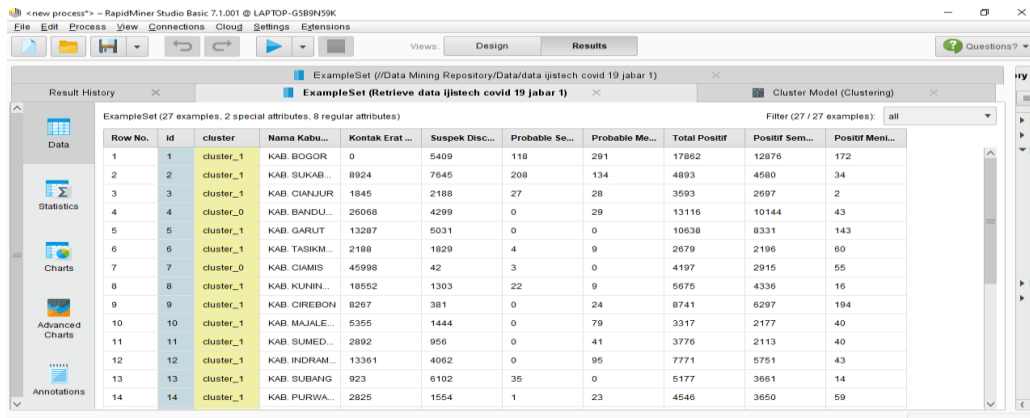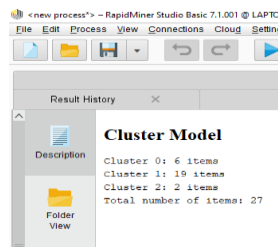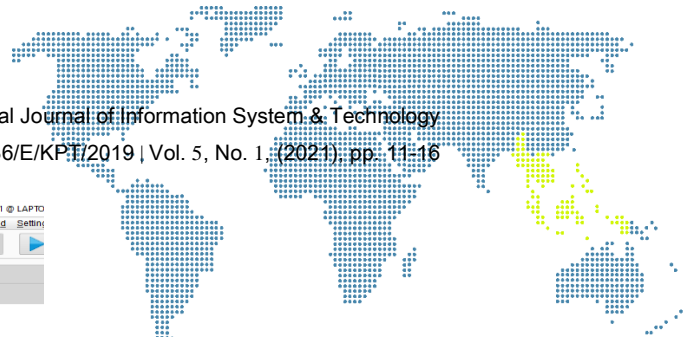**Figure 1.** The process of reading data with Rapidminer

Figure 2. West Java Covid 19 Clustering Results
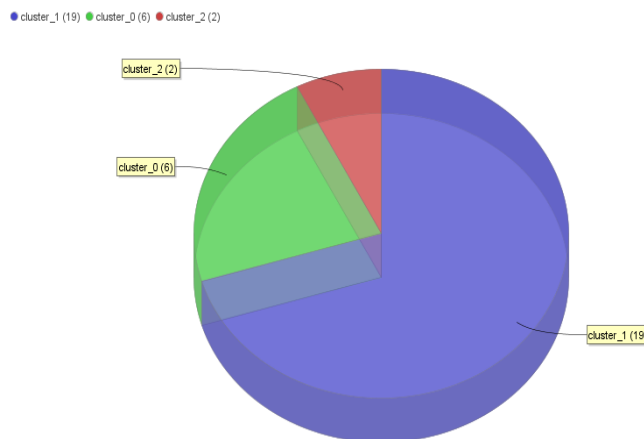

Figure 3. Clustering Results by province

Figure 4. Diagram of the parameters tested

4. Conclusion

Based on the research conducted as well as the implementation and testing that was done, it can be done grouping the Covid 19 data in West Java into 3 clusters, namely Cluster 0 with a high level of Covid 19 spread in districts/cities, namely Bandung City, Bogor, Ciamis Regency, Karawang, Bekasi, Bandung. Cluster 1 with moderate levels of covid spread in districts / cities, including Garut, Tasikmalaya, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Subang, Purwakarta, West Bandung, Pangandaran, Sukabumi City, Cirebon, Cimahi, Tasikmalaya, Banjar districts. Whereas for cluster 2 with a low level of distribution in the districts/cities of Bekasi and Depok.

References

[1]   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, pp. 52–59, 2020.

[2] A. H. A. R. S. R. Halfis, "Analisis Algoritma Klasifikasi C 4.5 Untuk Memprediksi Keberhasilan ImmunotherapyPada Penyakit Kutil," *J. Tek. Komput. AMIK BSI*, vol. v, no. N0.2, pp. 155–160, 2019.

[3] T. Khotimah and Darsin, "Clustering Perkembangan Kasus Covid-19 Di Indonesia Menggunakan Self Organizing Map," *J. Dialekt. Inform.*, vol. 1, no. 1, pp. 23–26, 2020.

[4] S. Hanoatubun, "Dampak Covid – 19 terhadap Prekonomian Indonesia," *EduPsyCouns J. Educ. Psychol. Couns.*, vol. 2, no. 1, pp. 146–153, 2020.

[5] A. Amalia and N. Sa'adah, "Dampak Wabah Covid-19 Terhadap Kegiatan Belajar Mengajar Di Indonesia," *J. Psikol.*, vol. 13, no. 2, pp. 214–225, 2020.

[6] B. S. A. U. A. Kafabih, "Sektor Pariwisata Indonesia Di Tengah Pandemi Covid 19," *J. Din. Ekon. Pembang.*, vol. 4, no. 1, pp. 383–389, 2021.

[7] P. J. Barat, "Pusat Informasi dan Koordinasi Covid 19 Provinsi Jawa Barat," 2021. [Online]. Available: https://pikobar.jabarprov.go.id.

[8] P. P. P. A. N. . Sukma Sindi, Weni Ratnasari Orktapia Ningse, Irma Agustika Sihombing and D. H. Fikrul Ilmi R.H.Zer, "Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 166–173, 2020.

[9] Samudi, S. Widodo, and H. Brawijaya, "The K-Medoids Clustering Method for Learning Applications during the COVID-19 Pandemic," *Sinkron*, vol. 5, no. 1, pp. 116–121, 2020.

[10] "Bagian 1: Kombinasi Metode Klastering dan Klasifikasi (Kasus Pandemi Covid-19 di Indonesia)," *Agus Perdana Wind. Ulfah Indriani Mokhamad Ramdhani Raharjo Linda Sari Dewi*, vol. 4, no. 3, pp. 855–862, 2020.

[11] I. Kamila, U. Khairunnisa, and Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, pp. 119–125, 2019.

[12] A. P. Windarto, "Penerapan Data Mining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means Clustering," *J. Teknol. Inf. Tecno.com*, vol. vol.16, no. no.4, pp. 348–357, 2017.

[13] R. R. Aria and S. Susilowati, "Analisa Data Penjualan SaRa Collectionmenggunakan metode Apriori," *J. Tek. KomputerAMIK BSI*, vol. 7, no. 1, pp. 68–73, 2021.

[14] E. Irawan, S. P. Siregar, I. S. Damanik, and I. S. Saragih, "Implementasi Algoritma K-Medoids untuk Pengelompokkan Sebaran Mahasiswa Baru," *J. Ris. Sist. Inf. Dan Tek. Inform.*, vol. 5, no. 2, pp. 275–281, 2020.