

FINAL

by Ronald Sebastian

Submission date: 04-Aug-2022 09:05AM (UTC-0500)

Submission ID: 1878780071

File name: Java_Province,_Indonesia_Using_the_K-Means_Algorithm-Final.docx (589.89K)

Word count: 3866

Character count: 21597

Application of Data Mining to Group the Spread of Covid-19 in West Java Province, Indonesia Using the K-Means Algorithm

¹Ronald Sebastian, ²Christina Juliane (10 pt)

^{1,2} Department of Business Information Systems, Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI (9 pt)
Email: ¹ronaldsebastian58@gmail.com, ²christina.juliane@likmi.ac.id

Article Info

Article history:

Received May 12th, 2019

Revised Jun 20th, 2020

Accepted Jul 26th, 2020

Keyword:

Data Mining

Covid-19

K-Means

Clustering

ABSTRACT (10 PT)

Covid-19 cases in Indonesia have not subsided. The spread of COVID-19 cases has reached provinces in Indonesia such as West Java, which is one of the many locations where the virus has been detected. COVID-19 cases have spread to 28 districts and cities in West Java. Researchers must determine the level of distribution of COVID-19 cases which are divided into three clusters, namely high, medium, and low clusters, so that the West Java Regional Government can take action in an effort to prevent the spread of COVID-19 cases. Researchers use data mining and the K-means Clustering algorithm to examine the distribution of COVID-19 cases. This data set for the study of the spread of COVID-19 in West Java Province, covers data for the period August 1, 2020 to July 15, 2022. To perform K-means Clustering on the data set, researchers used RapidMiner Studio 9.10. The results of this study indicate that in West Java there are two cities with the highest Covid-19 clusters, namely Bekasi and Depok, six cities and district in the medium cluster, namely city of Bogor, Bandung, and Karawang District, Bekasi, Bandung and Bogor, and The twenty district/cities in the lowest cluster for the spread of COVID-19 cases are the cities of Banjar, Cimahi, as well as the districts of West Bandung, Ciamis, Cianjur, Cirebon, Garut, Indramayu, Kuningan, Majalengka, Pangandaran, Purwakarta, Subang, Sukabumi, Sumedang, Tasikmalaya.

Copyright © 2022 Puzzle Research Data Technology

Corresponding Author: (10 pt)

Third Author,

Departement of Electrical and Computer Engineering,

National Chung Cheng University,

168 University Road, Minhsiung Township, Chiayi County 62102, Taiwan, ROC.

Email: thirdauthor@uin-suska.ac.id

DOI: <http://dx.doi.org/10.24014/ijaidm.v2i2.xxxx> (10 pt)

1. INTRODUCTION (10 PT)

The World Health Organization (WHO) officially declared the corona virus (COVID-19) as a pandemic on March 11, 2020 [1]. This means that COVID-19 has spread widely throughout the world. According to Worldometers data dated today (18/7/2022) there were 567,954,752 cases of COVID-19, which caused 6,388,035 deaths worldwide. Indonesia alone recorded 6,138,346 cases causing 156,859 deaths [2]. It is undeniable that COVID-19 cases continue to increase due to the unpreparedness of medical facilities and services for the surge in patients affected by COVID-19.

Covid-19 has a very fast pattern of transmission. The mode of transmission of this virus is by spreading through splashes of saliva when coughing, sneezing, even when exhaling produced by an infected person. Transmission can also occur when we breathe air, touch contaminated objects and the virus is spread when we touch our eyes, nose, or mouth [3].

The Indonesian government has also decided to take decisive action to prevent and reduce the spread of the Covid-19 virus as a result of its rapid spread throughout the world. There are several other actions taken by the Indonesian government, including regional quarantine, health protocols such as 5M (wash hands, stay away

from crowds, and limit mobility), and the temporary closure of physical activities in the workplace (such as factories, places of worship, and schools), to reduce the number of people contaminated by the virus [4].

Many studies have classified Covid-19 cases around the world [5] and in many countries [6], but a more in-depth investigation is needed for cases of the spread of Covid-19 in Indonesia [7], especially in West Java Province [3]. This study categorizes the spread of the Corona Virus in West Java Province based on the number of positive cases. Clustering is done using the K-means approach, which divides the data into high, medium, and low clusters depending on their similarity. [8]

Clustering means bringing together similar things, such as notes, observations, or seeing something similar and grouping them together. A cluster is a group of records that are similar to each other but different from other cluster records. Clustering tries to divide the entire data set into groups that have a lot in common. The similarity between records in one group is given the most weight, while the similarity between records in another group is given the least weight. Clustering is used in data mining to find patterns in a scattered set of data, which can help with data analysis. Similarity of objects is usually obtained from the proximity of attribute values that describe data objects, while data objects are usually displayed as points in multidimensional space [9].

K-Means is a non-hierarchical data clustering algorithm that tries to partition existing data in the form of one or more clusters/groups. In this algorithm, groups of data with the same characteristics are placed together in one cluster, while data with different characteristics are placed in another cluster [9]. The K-Means algorithm was used in this study to classify the spread of the Covid-19 virus. It was decided to use this algorithm for clustering because it has been thoroughly tested and is often used by researchers.

Several studies have used the k-means clustering algorithm for various purposes, including Mapping of Prospective Jamkesda Recipients [10], Grouping Tourist Visits to Featured Tourist Attractions in Prov. DKI Jakarta [11], Determining Candidates for Bidik Misi Scholarship Recipients in Polbeng [12], Clustering in Determining Superior Class Students [13]. Table 1 presents a summary of several recent studies that utilize the K-Means clustering algorithm.

Table 1. Research Related to Data Clustering Using the K-Means Algorithm

Paper	Research purposes	Research result
[10]	This study aims to analyze the performance between the FCM algorithm and K-means which is implemented on poverty data in Girijati Village, Purwosari into 3 clusters.	The application of the K-Means Method for community mapping based on the poverty level of the Kemuning sub-district was successfully applied to the application so that the recipients of JAMKESDA assistance were right on target.
[11]	The purpose of this study is to recommend a transaction history data that is processed by the Content-based method and the Collaborative method, then a hybrid process is carried out and produces the recommended tour package recommendations for tourists.	Grouping the number of tourists to the top attractions in Prov. DKI Jakarta can be applied with the K-Means clustering method.
[12]	The purpose of this study is to develop a web-based information system to manage centralized assessment data in presenting reports on student learning outcomes and grouping students in superior classes by implementing the K-means Clustering algorithm.	Based on the results of the discussion that has been described, the conclusion at the research stage is that the implementation of the k-means clustering algorithm into the clustering information system provides the results of an effective data grouping classification and the process of each iteration of the centroid distance rotation, the determination of the cluster point is formed, student data as a reference object saves more time to cluster the superior class.
[13]	The purpose of this study is to determine the clustering of scholarship applicants so that they can provide recommendations that are very worthy, worthy, with consideration and less worthy to receive scholarships with 6 criteria.	Based on the research that has been done, the K-Means algorithm is able to group prospective bidik misi scholarship recipients into 4 clusters, where cluster 0 means giving recommendations with consideration, cluster 1 provides very feasible recommendations, cluster 2 provides appropriate recommendations and cluster 3 provides poor recommendations. worthy.

The application of the K-Means Algorithm in this study using Microsoft Excel and RapidMiner applications aims to group data on Covid-19 cases. In simple terms Rapid Miner is an application created to

process data into more useful and useful information. Rapid Miner is open source software. Rapid Miner is a standalone software for data analysis and data mining engine that can be used in conjunction with its own product. All versions of Rapid Miner are compatible because they are written in Java [9], [14].

The collection of study data on the spread of COVID-19 in West Java Province includes data for the period August 1, 2020 to July 15, 2022. It is hoped that using the Corona virus distribution data that has been collected using the K-means Clustering Algorithm will help decision makers reduce the spread of the Corona virus and number of positive COVID-19 patients.

2. RESEARCH METHOD

In this study, the method used is the data mining method as follows. (a). Data collection stage; The data used in this study was collected from the official website of the West Java Covid-19 Information & Coordination Center with the url address <https://pikobar.jabarprov.go.id>. The data used for this study comes from monitoring the transmission of the corona virus from 1 August 2020 to 15 July 2022, (b). Data processing stage; The data that has been obtained will be processed first to be clustered. In the previous stage, the data for each city/district will be summed for each case so that at this stage the calculation of the value will be processed at the clustering stage, (c). Clustering Stage; clustering is an unattended form of categorization in which a set of data objects is automatically partitioned into several classes. This can be achieved using a number of distance algorithmic formulas and procedures, particularly those relating to Euclidean Distance. The purpose of cluster analysis is to partition the data set into groups with similar characteristics. In determining clusters based on available data, a flowchart is needed to make it easier to determine the calculation flow to find the results of applying clusters to the data to be processed [15]. Figure 1 below is a flowchart in determining clusters with K-Means.

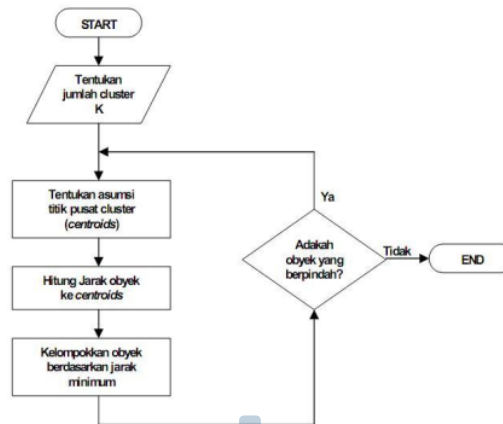


Figure 1. Flowchart of K-Means Algorithm Stages

At this stage the cluster is divided into 3 according to the level of spread of covid-19, including: high cluster, medium cluster, low cluster, and (d). Analysis Phase; At this stage, data analysis is carried out on the number of levels of the spread of COVID-19 in cities/districts in West Java Province. The data obtained were processed using the calculation of the weights of each index. In the previous stage, it was determined that it would be clustered into 3 clusters, namely clusters with high, medium, and low levels of covid spread. At this stage, the results will be analyzed [11].

2.1. Data Mining

By definition, data mining is the extraction and identification of relevant information from large data sets using statistical techniques, mathematics, artificial intelligence, and machine learning techniques (Silitonga, 2016). By using data from databases, data mining is able to extract important new information from it (Nur, Zarlis and Nasution, 2017). The ability to uncover the desired pattern in a large database is another advantage of data mining, which is used to assist future decision making (Pramesti, Furqon and Dewi, 2017). Descriptive and predictive data mining are two subcategories of data mining (Metisen and Sari, 2015). Data can be mined in two ways: descriptive and predictive. Descriptive mining is used to identify key properties in a data set, whereas predictive mining seeks to forecast future patterns in the data (Metisen and Sari, 2015). Classification, estimation, prediction, and clustering are some of the methods used in data mining (Supriyadi et al., 2018).

Data mining techniques, such as market basket analysis, have been used to change the structure of both modern and old markets (Soetopo, Tannady and Nurprihatin, 2017),[7].

2.2. Clustering Analysis

Multivariate statistical techniques, such as clustering, are used to express groupings of observations (Klimberg, Ratick and Smith, 2017). With clustering, we can organize data objects into small groups that are more closely related to each other, thereby maximizing the use of the data we collect (Kumar and Toshniwal, 2016). Clustering algorithms can be used to identify dense areas, find distribution patterns, and determine the relationship between each piece of data (Metisen and Sari, 2015),[7].

Cluster analysis uses hierarchical and non-hierarchical methodologies (Silvi, 2018). 1). Hierarchical approach by starting with two or more things that are most similar and moving on to the things that are closest to the second, this strategy creates a hierarchy or level between them (Goreti Nova and Wahyuningsih, 2016). Non-Hierarchical Approach This technique initially determines how many clusters there are in the data (Goreti, Nova and Wahyuningsih, 2016),[7].

There are several advantages of cluster analysis (Talakua, Leleury, and Taluta, 2017). 1. Able to classify large and massive collections of observational data. The analysis procedure will be easier if the data is broken down into groups. Useful for ordinal, interval, and ratio data. [7].

2.3. K-Means Algorithm

K-means, a non-hierarchical data grouping approach, can be used to divide data into two or more categories. Data groups with the same characteristics are grouped together, while data groups with different characteristics are separated. This data must be grouped to meet the objectives of the clustering function, namely minimizing group variance and maximizing group variation [11].

3. RESULTS AND ANALYSIS

The purpose of this study is to use K-Means Clustering to group data on the distribution of COVID-19 cases in West Java. There are several steps that must be taken to perform k-means clustering, as follows:

3.1 Data analysis

This research relies on secondary data taken on the <https://pikobar.jabbarprov.go.id> website which lists the number of positive cases of Covid-19 in each of the 28 districts/cities in West Java as of July 15, 2022 [16].

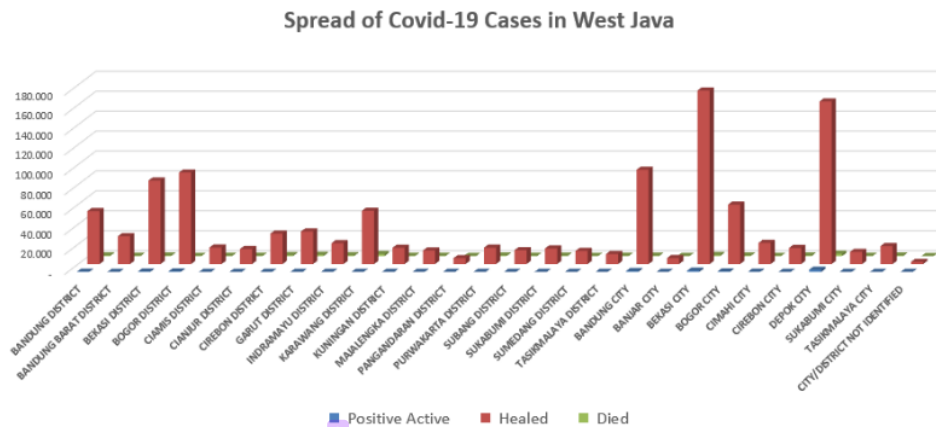


Figure 2. Distribution of Covid-19 Cases in West Java

3.2 Data processing

The stages of the K-Means Clustering algorithm process are as follows:

- Determine the amount of data to be in the cluster. High, medium and low are the three clusters that will be formed.
- Determine the center of the randomly generated initial centroid cluster (iteration 1). The following table shows the cluster center points to be calculated.

Table 2. Center Point Clusters of Each Beginning

Cluster	City/District	Positive Active	Healed	Died
C1	DEPOK CITY	2.223	163.206	2.252
C2	BOGOR DISTRICT	516	91.809	206
C3	PANGANDARAN DISTRICT	13	5.893	156

- c. Use the Euclidean Distance formula to determine the distance between each piece of data and the centroid. The data included in the cluster will be determined by the shortest distance between the data and the cluster at this stage.

$$de = \sqrt{(xi - si)^2 + (yi - ti)^2}$$

Information:

(x, y) = Object Coordinates

(s, t) = Centroid Coordinates

i = Number of objects

Table 3. Determining the Cluster Distance

No	Name of City/District	Positive Active	Healed	Died	C1	C2	C3	Shortest Distance	RESULTS
1	BANDUNG DISTRICT	229	53.135	634	110.100,949	38.677,433	47.244,912	38.677,433	C2
2	BANDUNG BARAT DISTRICT	90	27.901	299	135.335,904	63.909,487	22.008,599	22.008,599	C3
3	BEKASI DISTRICT	372	83.816	568	79.429,429	8.002,489	77.924,916	8.002,489	C2
4	BOGOR DISTRICT	516	91.809	206	71.446,704	0	85.917,487	0	C2
5	CIAMIS DISTRICT	23	16.568	328	146.667,122	75.242,714	10.676,390	10.676,390	C3
6	CIANJUR DISTRICT	27	15.042	227	148.194,109	76.768,560	9.149,286	9.149,286	C3
7	CIREBON DISTRICT	19	30.581	964	132.649,565	61.234,709	24.701,219	24.701,219	C3
8	GARUT DISTRICT	47	32.704	1.276	130.523,789	59.116,545	26.834,405	26.834,405	C3
9	INDRAMAYU DISTRICT	14	20.873	764	142.357,918	70.939,971	14.992,334	14.992,334	C3
10	KARAWANG DISTRICT	60	53.465	2.044	109.762,511	38.390,735	47.609,473	38.390,735	C2
11	KUNINGAN DISTRICT	18	16.312	286	146.923,703	75.498,685	10.419,812	10.419,812	C3
12	MAJALENGKA DISTRICT	11	13.602	267	149.633,519	78.208,654	7.709,799	7.709,799	C3
13	PANGANDARAN DISTRICT	13	5.893	156	157.342,484	85.917,487	0	0	C3
14	PURWAKARTA DISTRICT	22	16.386	606	146.845,722	75.425,678	10.502,649	10.502,649	C3
15	SUBANG DISTRICT	46	13.887	299	149.347,639	77.923,473	7.995,347	7.995,347	C3
16	SUKABUMI DISTRICT	35	15.639	356	147.595,398	76.171,666	9.748,077	9.748,077	C3
17	SUMEDANG DISTRICT	35	13.150	177	150.086,296	78.660,476	7.257,064	7.257,064	C3
18	TASIKMALAYA DISTRICT	17	10.017	317	153.217,102	81.793,597	4.127,143	4.127,143	C3
19	BANDUNG CITY	865	94.624	449	68.619,135	2.846,941	88.735,574	2.846,941	C2
20	BANJAR CITY	14	5.946	144	157.289,640	85.864,490	54,351	54,351	C3
21	BEKASI CITY	1.198	174.214	1.051	11.120,661	82.412,154	168.327,551	11.120,661	C1
22	BOGOR CITY	424	59.617	547	103.618,649	32.193,937	53.726,995	32.193,937	C2
23	CIMAHI CITY	176	21.174	200	142.061,571	70.635,819	15.281,933	15.281,933	C3
24	CIREBON CITY	64	16.175	568	147.056,493	75.636,217	10.290,377	10.290,377	C3
25	DEPOK CITY	2.223	163.206	2.252	0	71.446,704	157.342,484	0	C1
26	SUKABUMI CITY	13	12.171	290	151.063,910	79.639,633	6.279,430	6.279,430	C3

Title of manuscript is short and clear, implies research results... (First Author)

27	TASIKMALAYA CITY	68	17.973	593	145.258,461	73.838,373	12.088,027	12.088,027	C3
28	CITY/DISTRICT NOT IDENTIFIED	6	2.165	9	161.071,878	89.645,667	3.730,904	3.730,904	C3

Table 4. Results of Cluster Iteration 1

Cluster	City/District	Total
C1	21,25	2
C2	1,3,4,10,19,22	6
C3	2, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16 ,17, 18, 20, 23, 24, 26, 27, 28	20

- d. After all data is placed in the nearest cluster, the average number of members in the cluster is used to calculate the new center point. according to the following equation:

$$C = \frac{\sum m}{n}$$

Where:

C : centroid data

m : data member that belongs to a certain centroid

n : the amount of data that is a member of a certain centroid

- e. K-Means will iterate until the data grouping is the same as the previous iteration grouping.

In this case, there is no need to proceed to the second iteration because the results of the first iteration are identical. After that, the group is said to be convergent or already considered to be in its ideal state.

THE SPREAD OF COVID-19 IN WEST JAVA

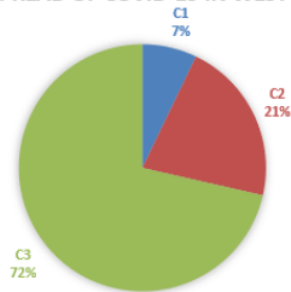


Figure 3: Clustering Iteration 1

3.3 RapidMiner's K-Means Clustering Process

Figure 3 shows how the k-means clustering method in RapidMiner is used to analyze the distribution of COVID-19 cases in West Java:

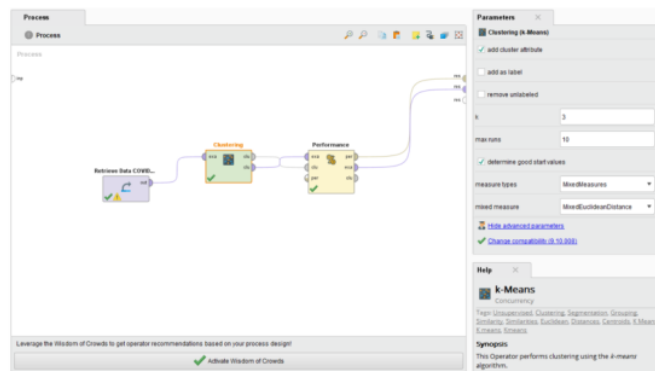


Figure 4. Clustering Process on RapidMiner

3.4 RapidMiner K-Means Clustering Process Results

The following are the results of the cluster model after the K-means Clustering process is carried out.

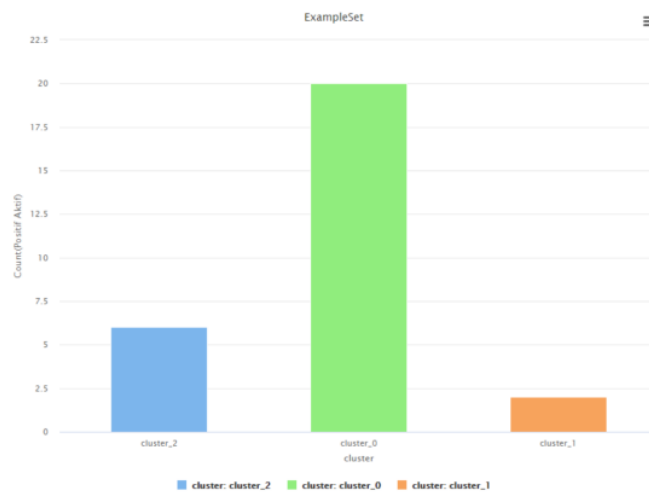


Figure 5. Clustering Results

The results of the k-means clustering process using rapidminer, for three cluster groups containing 28 data, can be seen in Figure 6:

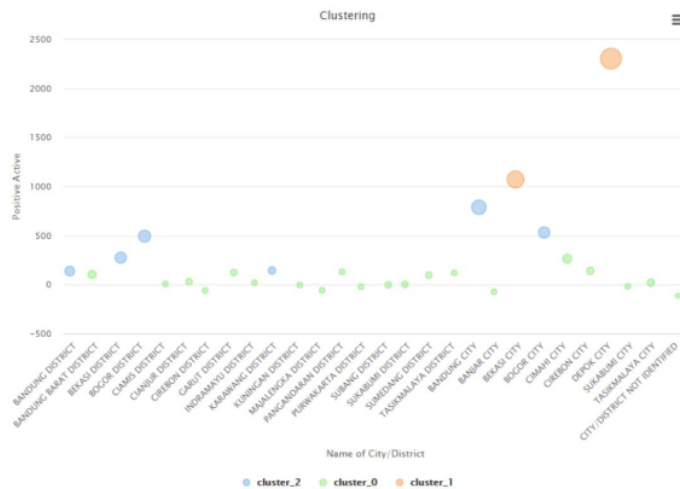


Figure 6. Final Result of Rapidminer Grouping

The results in Figures 5 and 6 are obtained in iteration 1 because iteration 1 is said to be identical and already in its ideal state, the results can be seen where cluster 1 on RapidMiner is a high cluster of 2 districts/cities which can be seen in the red dot, the medium cluster is cluster 2 totaling 6 district/cities on the green dot, while the low cluster, namely cluster 0, consists of 20 district/cities on the blue dot. Based on the explanation above regarding the stages of data processing and the results that have been displayed, it will be discussed regarding the relationship of the results obtained from RapidMiner and Microsoft Excel explaining that the results of the manual calculation of the k-means algorithm have the same value from the number of each cluster, namely for high cluster, medium cluster and low cluster.

4. CONCLUSION

The K-Means Clustering algorithm was used to analyze and test the data on the distribution of Covid-19 cases in West Java in the period 1 August 2020 to 15 July 2022 obtained from the <https://pikobar.jabarprov.go.id> site, and the results showed that the algorithm K-Means Clustering is very accurate. As a result, the findings of this study can be summarized as follows. Covid-19 cases in West Java were analyzed for distribution using the K-Means Clustering technique in Microsoft Excel and the Rapidminer software. Clusters are divided into 3, namely: high, medium, and low clusters. Based on a total of 28 district and city data in West Java, the cities of Bekasi and Depok are included in the highest cluster of COVID-19 cases. In addition, the cities of Bogor, Bandung, and the district of Karawang, Bekasi, Bandung and Bogor are included in the medium cluster. The lower cluster consists of the cities of Banjar, Cimahi, and the districts of West Bandung, Ciamis, Cianjur, Cirebon, Garut, Indramayu, Kuningan, Majalengka, Pangandaran, Purwakarta, Subang, Sukabumi, Sumedang, Tasikmalaya. The results obtained from the K-Means algorithm can be applied to RapidMiner with the same validation value and can be taken into consideration by the West Java Provincial Health Office to take further decisions regarding the policies implemented to handle the level of Covid-19 cases in districts and cities in Indonesia. West Java, so the quality of the handling is better.

5. REFERENCE

- [1] "WHO Resmi Sebut Virus Corona Covid-19 sebagai Pandemi Global Halaman all - Kompas.com." <https://www.kompas.com/sains/read/2020/03/12/083129823/who-resmi-sebut-virus-corona-covid-19-sebagai-pandemi-global?page=all> (accessed Jul. 27, 2022).
- [2] "Indonesia COVID - Coronavirus Statistics - Worldometer." <https://www.worldometers.info/coronavirus/country/indonesia/> (accessed Jul. 27, 2022).
- [3] Elsa Ramadanti and Muhamad Muslih, "ANALISIS PERSEBARAN KASUS COVID-19 DI JAWA BARAT MENGGUNAKAN METODE K-MEANS CLUSTERIN | Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra," *SISMAPIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika)*, 2021. <https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/41> (accessed Jul. 18, 2022).

- [4] "Protokol Kesehatan 5 M di Masa PPKM – Pusat Studi Lingkungan Hidup UGM." <https://pslh.ugm.ac.id/protokol-kesehatan-5-m-di-masa-ppkm/> (accessed Jul. 27, 2022).
- [5] V. Zariakas, S. G. Pouloupoulos, Z. Gareiou, and E. Zervas, "Clustering analysis of countries using the COVID-19 cases dataset," *Data in Brief*, vol. 31, Aug. 2020, doi: 10.1016/J.DIB.2020.105787.
- [6] M. Azarafza, M. Azarafza, and H. Akgün, "Clustering method for spread pattern analysis of coronavirus (COVID-19) infection in Iran," *Journal of Applied Science, Engineering, Technology, and Education*, vol. 3, no. 1, pp. 1–6, Apr. 2021, doi: 10.35877/454RI.ASCI31109.
- [7] G. D. Rembulan, T. Wijaya, D. Palullungan, K. N. Alfina, and M. Qurthuby, "Kebijakan Pemerintah Mengenai Coronavirus Disease (COVID-19) di Setiap Provinsi di Indonesia Berdasarkan Analisis Klaster," *JIEMS (Journal of Industrial Engineering and Management Systems)*, vol. 13, no. 2, pp. 74–86, Sep. 2020, doi: 10.30813/JIEMS.V13I2.2280.
- [8] A. Solichin and K. Khairunnisa, "Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means," *Fountain of Informatics Journal*, vol. 5, no. 2, pp. 52–59, Oct. 2020, doi: 10.21111/FIJ.V5I2.4905.
- [9] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS," *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [10] M. N. V. Waworuntu and M. F. Amin, "PENERAPAN METODE K-MEANS UNTUK PEMETAAN CALON PENERIMA JAMKESDA," *KLIK - KUMPULAN JURNAL ILMU KOMPUTER*, vol. 5, no. 2, pp. 190–200, Sep. 2018, doi: 10.20527/KLIK.V5I2.157.
- [11] L. Maulida *et al.*, "Penerapan Data Mining dalam Mengelompokkan Kunjungan Wisatawan ke Objek Wisata Unggulan di Prov. DKI Jakarta dengan K-Means," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 2, no. 3, pp. 167–174, Mar. 2018, doi: 10.14421/jiska.2018.23-06.
- [12] F. Profesio Putra, P. Negeri Bengkalis, J. Bathin Alam, and S. Alam, "K-MEANS UNTUK MENENTUKAN CALON PENERIMA BEASISWA BIDIK MISI DI POLBENG," *INOVTEK Polbeng - Seri Informatika*, vol. 1, no. 1, pp. 87–94, Jun. 2016, doi: 10.35314/ISI.V1I1.129.
- [13] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustring dalam Penentuan Siswa Kelas Unggulan," *Jurnal Tekno Kompak*, vol. 15, no. 2, pp. 25–36, Aug. 2021, doi: 10.33365/JTK.V15I2.1162.
- [14] 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," *Jurnal Ilmu Komputer dan Informatika*, vol. 1, no. 2, pp. 109–132, Nov. 2021, doi: 10.54082/JIKI.13.
- [15] R. W. Sari, A. Wanto, and A. P. Windarto, "IMPLEMENTASI RAPIDMINER DENGAN METODE K-MEANS (STUDY KASUS: IMUNISASI CAMPAK PADA BALITA BERDASARKAN PROVINSI)," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 2, no. 1, Oct. 2018, doi: 10.30865/KOMIK.V2I1.930.
- [16] "Dashboard Statistik Kasus COVID-19 Provinsi Jawa Barat - Dashboard Jabar." <https://dashboard.jabarprov.go.id/id/dashboard-pikobar/trace/statistik> (accessed Jul. 27, 2022).

FINAL

ORIGINALITY REPORT

19%

SIMILARITY INDEX

12%

INTERNET SOURCES

11%

PUBLICATIONS

6%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Institut Pertanian Bogor

Student Paper

4%

2

ijair.id

Internet Source

2%

3

Nurul Hidayati. "APPLICATION OF CLUSTERING ANALYSIS TO DATA DISTRIBUTION OF COVID-19 IN BENGKULU PROVINCE", BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2022

Publication

2%

4

M W F Umam, M Fatekurohman, D Anggraeni. "Hybrid clustering and classification methods to find out the pattern of the spread of covid-19 in East Java province", Journal of Physics: Conference Series, 2022

Publication

2%

5

media.neliti.com

Internet Source

2%

6

Rusdiansyah Rusdiansyah, Harun Al Rasyid, Suryanto Sosrowidigdo. "Implementation of

1%

address recording management using the K-Means clustering classification algorithm in Kebayoran District, DKI Jakarta", SinkrOn, 2021

Publication

7

Kristin D R Sianipar, Septri Wanti Siahaan, Indra Gunawan. "Penerapan Algoritma K-Means dalam Pengelompokan Rumah Tangga Menurut Provinsi dan Status Kepemilikan Rumah Milik Sendiri", SinkrOn, 2021

Publication

1 %

8

download.garuda.kemdikbud.go.id

Internet Source

1 %

9

mail.ijair.id

Internet Source

1 %

10

www.openjournal.unpam.ac.id

Internet Source

1 %

11

id.wikipedia.org

Internet Source

1 %

12

Jaroji Jaroji, Danuri Danuri, Fajri Profesio Putra. "K-MEANS UNTUK MENENTUKAN CALON PENERIMA BEASISWA BIDIK MISI DI POLBENG", INOVTEK Polbeng - Seri Informatika, 2016

Publication

1 %

13

Internet Source

1 %

14

Agus Perdana Windarto. "Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method", International Journal of Artificial Intelligence Research, 2017
Publication

1 %

15

ejournal.nusamandiri.ac.id
Internet Source

1 %

16

www.semanticscholar.org
Internet Source

1 %

Exclude quotes On

Exclude matches < 1 %

Exclude bibliography On