CLUSTERING OF COVID-19 VACCINATION RECIPIENTS IN DKI JAKARTA USING THE K-MEDOIDS ALGORITHM

Frista Millenia Trisuciana[1]
School of Industrial and System Engineering
Telkom University
Bandung, Indonesia
fristamilleniat@student.telkomuniversity.ac.id

Deden Witarsyah[2]
School of Industrial and System Engineering
Telkom University
Bandung, Indonesia
dedenw@telkomuniversity.ac.id

Edi Sutoyo[3]
School of Industrial and System Engineering
Telkom University
Bandung, Indonesia
edisutoyo@telkomuniversity.ac.id

José Manuel Machado[4]
School of Engineering
University of Minho
Portugal
jmac@di.uminho.pt

Abstract— The COVID-19 pandemic has become a serious problem that has attacked various aspects of life such as social, economic, religious, and others. The government has held a COVID-19 vaccination program as an effort to deal with the COVID-19 problem since January 13, 2021. Many problems occurred due to difficulties in dividing the vaccination recipient areas. This is due to the large number of regions with different conditions for each region. One of the efforts to assist the process of processing large vaccination data is data mining techniques and using the clustering method with the K-medoids algorithm. In this study, data on COVID-19 vaccination was grouped in the East Jakarta area using the K-medoids algorithm clustering method. The calculation is carried out using the Euclidean Distance equation and the value of S > 0. The grouped area categories are at the kelurahan level which will then be divided into several clusters. The clustering process was carried out with RapidMiner on 267 kelurahan data on four main attributes, namely the number of targets, the number of vaccine doses 1, the number of vaccine doses 2, and the number of vaccine doses 3. The clustering process was carried out in 6 simulations with variations of k medoids as much as 2 to 7. The results of clustering show the best number of clusters obtained in the simulation is cluster 6 with the smallest Davies Bouldin Index (DBI) value of 0.209. The clusters obtained are clusters 0 to cluster 5. The cluster that is prioritized in giving vaccinations is cluster 2 with 67 items because its members are areas in DKI Jakarta and give a high score in cases of COVID-19 compared to other clusters.

Keywords— Pandemic, COVID-19, Vaccination, Clustering, K-Medoids

I. INTRODUCTION

The widespread spread of diseases caused by the COVID-19 virus which was designated as a pandemic by WHO on March 12, 2020. As a result of this COVID-19 virus, many patients died [1]. Currently, experts and scientists around the world are trying to do research to find a COVID-19 vaccine [2]. After experts and scientists tried to find a vaccine, finally a COVID-19 vaccine has been found and the government has held a COVID-19 vaccination program in the community. The public can see who has received the vaccine by age group and region through the Open Government Data (OGD) provided by the government.

Open government data is an open government doctrine regarding the data it has for certain things so that the public can freely use the data. Vaccination file history can be made in the form of Open Government Data which has now been implemented by major cities in Indonesia, one of which is DKI Jakarta. DKI Jakarta has launched an open data portal. Vaccination file history is divided into two bases, the first based on the ward and sub-district and the second based on the RT which can be accessed at DKI Jakarta vaccination page[1]. One of the efforts to assist the process of processing large vaccination data is data mining techniques.

Technically, data mining can be referred to as the process of finding correlations or patterns from hundreds or thousands of fields from a large relational database [3]. Data mining can be used to explore added value in the form of information obtained by recognizing important patterns from the data contained in the database [4]. Clustering is one of the processes in data mining that can be used for grouping data on areas (ward) of vaccination recipients in DKI Jakarta [5].

Clustering is a process of grouping records, observations, or grouping classes that have the same object [6]. Implementation of clustering can be done by applying algorithms, one of which is the K-Medoids algorithm [7]. The K-medoids algorithm method is used in making a cluster system to group areas affected by COVID-19 so as to provide mapping results in the form of clusters of affected areas or villages and knowledge can be obtained in the form of priority information on areas that need to be vaccinated [8].

In this study, researchers used the Clustering technique to group data on areas (ward) of vaccination recipients in DKI Jakarta. The K-Medoids Algorithm method is used in making a cluster system to group areas affected by COVID-19 so as to provide mapping results in the form of clusters of affected areas or villages and knowledge can be obtained in the form of priority information on areas that need to be vaccinated [9]. K-Medoids is one of the partitioning methods, because it uses the most centralized objects (medoids) in the cluster to be the cluster center of the average value of objects in a cluster. The K-medoids method is more suitable for grouping data than the K-Means method [10]. This research can be a solution to make it easier to find out which areas are most common in people who have not vaccinated against COVID-19, so that the government can increase the dose of vaccine in areas that still receive few COVID-19 vaccinations.

II. BASIC THEORY

A. Data Mining

Data mining is a set of heuristics and calculations that form a

---

data model that becomes a challenge in choosing the right or most appropriate algorithm to be used in solving a particular problem. Some algorithms can perform a clustering process that can group one or more discrete variables based on other attributes in data collection [11]. Data mining techniques are used to examine large databases as a way to find new and versatile patterns. Data mining is usually used to improve future decision making based on information obtained from the past, for example for prediction, estimation, association, clustering, and description [5].

B. Clustering

Clustering is a data mining technique that aims to group data based on the information found in the data. The grouping is based on the similarity between the data so that the data in the same cluster is homogeneous. Clustering is a popular partitioning algorithm known as Medoid (PAM) as k-medoid clustering. Clustering is general unattended machine learning where data sets must be automatically partitioned into "clusters" to make objects in the same cluster more similar and objects in different clusters more distinct [12]. Clustering is seen as one way of mapping COVID-19 cases to deal with the pandemic. Over the past few years, many variants of clustering methods have been proposed. Broadly speaking, each of these grouping methods is seen as a hierarchical or non-hierarchical category. The hierarchical method is used to group observations structurally based on their similarity, the number of desired clusters isnt predetermined. While the non-hierarchical method, on the other hand, starts with determining the desired number of clusters [13].

C. Algoritma K-Medoids

K-Medoids is a classical partitioning clustering technique that groups a data set of n objects into k groups known as a priori. K-medoids can be interpreted as cluster objects whose average difference for all objects in the cluster is minimal, i.e. the most located point in the cluster [14]. K-Medoids is one of the partitioning methods because it uses the most centralized object (medoids) where the cluster becomes the center of the cluster of the average value of objects in a cluster. The K-medoids method is more suitable for grouping data than the method K-Means [10]. The K-Medoids algorithm is one of the strong clustering algorithms in dealing with an outlier problem. K-Medoids perform better when the data set ranges from small to medium. When the size of the data set is large, the computational complexity becomes higher and makes it unsuitable [15].

D. Davies Bouldin Index

Davies Bouldin Index Davies Bouldin Index to develop clusters in general based on the quantity and proximity between cluster members. Calculation of the Davies Bouldin Index value based on the comparison of the ratio of the i-th cluster and the j-th cluster [9]. The smaller the Davies Bouldin Index value, the better the resulting cluster. The calculation of the DBI value is presented in the equation [16]. The Davies-Bouldin Index (DBI) is one of the methods of internal validity in evaluating a cluster, Internal Validity is what DBI does for how well the cluster has been done by calculating the quantity and derived features of the data set [17]. The steps to calculate the Davies Bouldin Index (DBI) are as follows

$$SSB = \frac{2}{M(M-1)} \sum_{I=1}^{M} \sum_{J=1, J \neq 1}^{M} || C_I - C_J ||^2$$

$$SSB_{i,j} = d (c_i, c_j)$$

After obtaining the value of the inter distance and intra distance, a ratio measurement (Rij) will be carried out to produce a comparison value between the i-th cluster and the j.3 cluster.

$$R_{I,J} = \frac{SSW_i + SSW_J}{SSB_{I,J}}$$

The ratio value generated from the equation will be used to find the DBI value using the following equation.

$$DBI = \frac{1}{K} \sum_{i=1}^{K} max_{i \neq 1}(R_{i,j})$$

From these calculations, the DBI value will be generated, the smaller the DBI value or the closer to 0 the better the cluster obtained [18].

E. Rapid Miner

Rapid Miner is a software developed by dr. Markus Hofmann from Blanchardstown Institute of Technology and Ralf Klinkenberg from rapid-i.com with a GUI (Graphical User Interface) display to make it easier for users to use this software. This software is open source and developed using Java programs under the GNU public license, and Rapid Miner can be run on any operating system. When using Rapid Miner, no special coding skills are needed because all the facilities are provided. Rapid Miner is dedicated to the use of data mining. The models provided are also quite numerous and complete, such as Bayesian models, modeling, tree induction, neural networks and others. Rapid Miner offers many methods from classification, grouping, association and much more [15].

III. IMPLEMENTATION AND DISCUSSION

A. Conceptual Model

A conceptual model is a set of diagrams related to certain factors that can have an impact or direction on a target or target condition. The conceptual model is related to the theories related to the research variables you want to study, namely the independent variable and the dependent variable [19]. Figure 1 is the conceptual model used in this study.
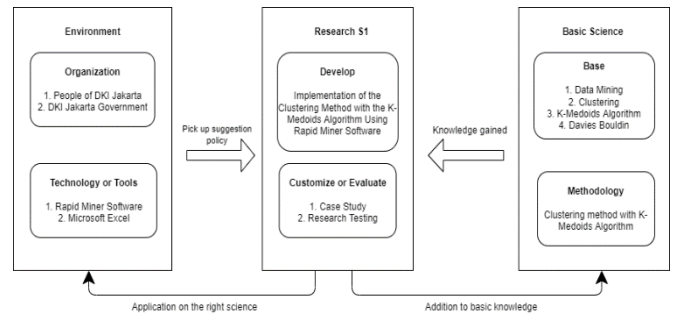

Figure 1. *Conceptual Model*

From Figure III.1 it can be concluded that the above modeling is based on the problems that exist in DKI Jakarta, namely determining the status of the most vaccination recipients in an area. The actors involved in this community are the people and the government of DKI Jakarta. This study uses the concept of data mining techniques using rapid mining software. The DKI Jakarta government can make decisions based on projected results in the form of clustering the status of COVID-19 vaccination recipients.

B. Vaccination Research Data

In this study, the data that will be studied by researchers is regarding data on the delivery of the COVID-19 vaccine for the DKI Jakarta area. The source of the research data was obtained from the DKI Jakarta Vaccination File History[2]. The data is taken in the form of excel data spreadsheets from June 20, 2021 to April 19, 2022. The data will be presented in the form of monthly averages during the normalization process. The following is one of the data presented on April 19, 2022 which is grouped by ward.

C. Data Selection

The research data that has been obtained is then selected or filtered first and several attributes are taken from the table for analysis. Attribute selection aims to take attributes that have a strong correlation with the characteristics of vaccine administration for an area. The attributes that will be used in this study are shown in Table 1 as follows.

Table 1. Data Selection

---

[2] DKI Jakarta vaccination page: https://riwayat-file-vaksinasi-dki-jakarta-jakartagis.hub.arcgis.com/

Attribute Name	Set Value
Ward	Label
Target	Float
Number of Doses 1	Float
Number of Doses 2	Float
Number of Doses 3	Float

D. Pre-Processing

After selecting the attributes and set values that will be used in clustering, the next step is pro-processing. Pre-processing, here we calculate the monthly average value of the dataset attributes used. The dataset used is from June 20, 2021 to April 19, 2022 or 11 months. The average calculation uses the number of doses of 1 vaccine for 1 month plus the number of vaccines given in the next 1 month divided by 11 months. The results of pre-processing with an average monthly value are as shown in Table 2 below.

Table 2. *Pre-Processing* Data

No	Ward	Target	Number of Dosis 1	Number of Dosis 2	Number of Dosis 3
1	Ancol	2177	1771	1545	538
2	Angke	2671	2205	1917	743
3	Bale Kambang	2643	2123	1841	520
4	Bali Mester	887	741	670	265
5	Bambu Apus	2390	2037	1857	669
6	Bangka	1961	1612	1448	463
7	Baru	2171	1883	1734	789
8	Batu Ampar	4354	3530	3104	989
9	Bendungan Hilir	1954	1623	1480	653
10	Bidara Cina	3335	2680	2380	809
…	………	….	…	…	…
267	Wijaya Kusuma	3539	2939	2665	1179

E. Data Normalization

After getting the research data, the next process is to normalize the data. The normalization process is also called the binning process, which is a transformation process to change the data value. Normalization is used to equalize the scale of data attributes into a smaller specific range such as -1 to 1 or 0 to 1. Min-Max Normalization is a normalization technique by performing linear transformations on the original data attributes to produce the same range of values. Min-Max Normalization maps a value v from attribute A to v' into the range [new_minA, new_MaxA]. Before the normalization process uses the min-max equation, the min-max value of each attribute is determined first. The min max values of each attribute are shown in Table 3 with the normalized data with the min max concept are shown in Table 4.

Table 3. *Min Max Value*

Attribut Name	Min Value	Max Value
Target	176	12224
Number of Doses 1	164	9873
Number of Doses 2	150	8598
Number of Doses 3	71	3060

Table 4. Data *Normalization*

No	Ward	v'Target	v'Number of Dosis 1	v'Number of Dosis 2	v'Number of Dosis 3
1	Ancol	0.1661	0.1655	0.1651	0.1563
2	Angke	0.0344	0.0348	0.0345	0.0352
3	Bale Kambang	0.0340	0.0334	0.0330	0.0235
4	Bali Mester	0.0098	0.0098	0.0102	0.0102
5	Bambu Apus	0.0305	0.0319	0.0334	0.0313
6	Bangka	0.0246	0.0247	0.0254	0.0205
7	Baru	0.0275	0.0293	0.0309	0.0375
8	Batu Ampar	0.0576	0.0574	0.0577	0.0480
9	Bendungan Hilir	0.0245	0.0249	0.0260	0.0304
10	Bidara Cina	0.0435	0.0429	0.0436	0.0386
…	………	….	…	…	…
267	Wijaya Kusuma	0.0464	0.0473	0.0492	0.0580

F. K-Medoids Process

K-medoids is one of the partitioning methods, because it uses the most centralized object (medoids) in the cluster to be the cluster center of the average object value in a cluster. The flow in conducting the data clustering process with the K-medoids algorithm is as follows.

1. Initialize k cluster centers (number of clusters)

2. Enter each data (object) into the nearest cluster using the euclidian Distance measure equation, with the following equation.

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_m - y_m)^2}$$

3. Randomly select an object in each as a candidate for new medoids.

4. Calculate the distance of each object that is in each of the new candidate medoids.

5. Calculate the total deviation (S) by calculating the value of new total_distance - old total_distance. If S < 0, then swap objects with cluster data to form a new set of k objects as medoids.

6. Repeat steps 3 to 5 until there is no change in the medoids, so that clusters and their respective cluster members are obtained.

The flow in conducting the data clustering process with the K-Medoids algorithm is as follows [20].
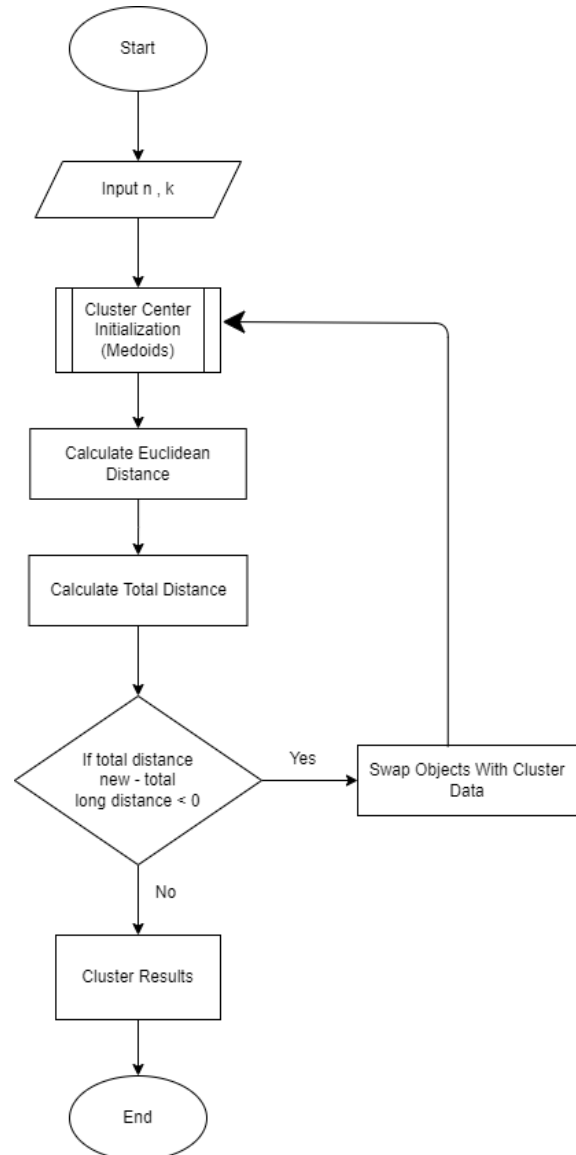

Figure 2. *K-Medoids Algorithm Calculation Flow*

Based on the process flow shown in Figure 2, it is possible to calculate manually the K-Medoids clustering using the normalized data in Table 6. The following is a manual cluttering calculation process with K-Medoids :

A. 1st Iteration Calculation Results

In each iteration random attribute will be determined, random attribute selection in iteration 1 is shown in Table 5. Based on the generated random attributes, it is obtained as in Table 6.

Table 5. Attribut 1st Iteration

Ward	v'Target	v'Number of Dosis 1	v'Number of Dosis 2	v'Number of Dosis 3
Warakas	0.0532	0.0559	0.0561	0.0441
Wijaya Kusuma	0.0464	0.0473	0.0492	0.0580

Table 6. Total Cost 1st Iteration

Ward	d_C1	d_C2	Proximity
Ancol	0.2218	0.2267	0.2218
Angke	0.0367	0.0322	0.0322
Bale Kambang	0.0428	0.0423	0.0423
Bali Mester	0.0853	0.0809	0.0809
Bambu Apus	0.0421	0.0381	0.0381
Bangka	0.0574	0.0543	0.0543
Baru	0.0452	0.0378	0.0378
Batu Ampar	0.0062	0.0200	0.0062
Bendungan Hilir	0.0537	0.0477	0.0477
Bidara Cina	0.0213	0.0208	0.0208
....	...	...	...
Wijaya Kusuma	0.0190	0.0000	0.0000
Total Cost (S)			12.3714

In the 1st iteration, the value of Total Cost (S) = 12.3714

The value of S will be compared with the next iteration until the difference between the new S and the old S is positive (S > 0).

B. 2nd Iteration Calculation Results

Next, random attribute selection is performed in iteration 2 which is shown in Table 7 with the results shown in Table 8.

Table 7. Attribut 2nd Iteration

Ward	v'Target	v'Number of Dosis 1	v'Number of Dosis 2	v'Number of Dosis 3
Ancol	0.1661	0.1655	0.1651	0.1563
Angke	0.0344	0.0348	0.0345	0.0352

Table 8. Total Cost 2nd Iteration

Ward	d_C1	d_C2	Proximity
Ancol	0.0000	0.2572	0.0000
Angke	0.2572	0.0000	0.0000
Bale Kambang	0.2645	0.0118	0.0118
Bali Mester	0.3066	0.0494	0.0494
Bambu Apus	0.2630	0.0063	0.0063
Bangka	0.2789	0.0223	0.0223
Baru	0.2643	0.0098	0.0098
Batu Ampar	0.2161	0.0418	0.0418
Bendungan Hilir	0.2739	0.0171	0.0171
Bidara Cina	0.2422	0.0156	0.0156
....	...	...	...
Wijaya Kusuma	0.2267	0.0322	0.0322
Total Cost (S)			10.1715

In the 2nd iteration, the value of Total Cost (S) = 10.1715
Then calculate the value of the deviation (S) with the equation
S = New Total Cost – Old Total Cost
S = 10.1715 – 12.3714
S = -2.1999
Because the value of S is not greater than 0 (S > 0), the calculation continues to iteration 3.

C. 3rd Iteration Calculation Results

The random attribute selection in iteration 3 is shown in Table 9 with the results in Table 10.

Table 9. Attribut 3rd Iteration

Ward	v'Target	v'Number of Dosis 1	v'Number of Dosis 2	v'Number of Dosis 3
Kramat	0.0337	0.0337	0.0318	0.0236
Kramat Jati	0.0403	0.0412	0.0416	0.0370

Table 10. Total Cost 3rd Iteration

Ward	d_C1	d_C2	Proximity
Ancol	0.2650	0.2464	0.2464
Angke	0.0119	0.0114	0.0114
Bale Kambang	0.0013	0.0189	0.0013

Bali Mester	0.0423	0.0602	0.0423
Bambu Apus	0.0086	0.0168	0.0086
Bangka	0.0147	0.0325	0.0147
Baru	0.0159	0.0205	0.0159
Batu Ampar	0.0489	0.0307	0.0307
Bendungan Hilir	0.0156	0.0283	0.0156
Bidara Cina	0.0233	0.0044	0.0044
....	...	...	...
Wijaya Kusuma	0.0427	0.0239	0.0239
Total Cost (S)			9.1315

In the 3rd iteration, the value of Total Cost (S) = 9.1315
Then calculate the value of the deviation (S) with the equation
S = New Total Cost – Old Total Cost
S = 9.1315 – 10.1715
S = -1.0399
S < 0 the calculation continues to iteration 4.

D. 4th Iteration Calculation Results

The random attribute selection in iteration 4 is shown in Table 11 with the results in Table 12.

Table 11. Attribut 4th Iteration

Ward	v'Target	v'Number of Dosis 1	v'Number of Dosis 2	v'Number of Dosis 3
Cakung Timur	0.0706	0.0728	0.0741	0.0677
Sungai Bambu	0.0350	0.0362	0.0371	0.0319

Table 12. Total Cost 4th Iteration

Ward	d_C1	d_C2	Proximity
Ancol	0.1840	0.2564	0.1840
Angke	0.0733	0.0044	0.0044
Bale Kambang	0.0808	0.0098	0.0098
Bali Mester	0.1227	0.0503	0.0503
Bambu Apus	0.0791	0.0073	0.0073
Bangka	0.0950	0.0226	0.0226
Baru	0.0807	0.0132	0.0132
Batu Ampar	0.0326	0.0405	0.0326
Bendungan Hilir	0.0901	0.0191	0.0191
Bidara Cina	0.0583	0.0143	0.0143
....	...	...	...
Wijaya Kusuma	0.0442	0.0328	0.0328
Total Cost (S)			7.6568

In the 4th iteration, the value of Total Cost (S) = 7.6568
Then calculate the value of the deviation (S) with the equation
S = New Total Cost – Old Total Cost
S = 7.6568 – 9.1315
S = -1.4748
The calculation continues to iteration 5.

E. 5th Iteration Calculation Results
The random attribute selection in iteration 5 is shown in Table 13 with the results in Table 14.

Table 13. Attribut 5th Iteration

Ward	v'Target	v'Number of Dosis 1	v'Number of Dosis 2	v'Number of Dosis 3
Karet Tengsin	0.0217	0.0222	0.0225	0.0178
Pondok Labu	0.0558	0.0594	0.0639	0.0682

Table 14. Total Cost 5th Iteration

Ward	d_C1	d_C2	Proximity
Ancol	0.2844	0.2035	0.2035
Angke	0.0276	0.0549	0.0276
Bale Kambang	0.0205	0.0640	0.0205
Bali Mester	0.0225	0.1040	0.0225
Bambu Apus	0.0217	0.0607	0.0217
Bangka	0.0055	0.0770	0.0055
Baru	0.0233	0.0610	0.0233
Batu Ampar	0.0684	0.0212	0.0212
Bendungan Hilir	0.0137	0.0709	0.0137
Bidara Cina	0.0422	0.0413	0.0413
....	...	...	...
Wijaya Kusuma	0.0597	0.0236	0.0236
Total Cost (S)			7.5260

In the 5th iteration, the value of Total Cost (S) = 7.5260
Next calculate the value of the deviation (S) with the equation
S = New Total Cost – Old Total Cost
S = 7.5260 - 7.6568

S = -0.1307
The calculation continues to iteration 6.

F. 6th Iteration Calculation Results
The random attribute selection in iteration 6 is shown in Table 15 with the results in Table 16.

Table 15. Attribut 6th Iteration

Ward	v'Target	v'Number of Dosis 1	v'Number of Dosis 2	v'Number of Dosis 3
Kemayoran	0.0235	0.0240	0.0245	0.0238
Taman Sari	0.0155	0.0161	0.0168	0.0187

Table 16. Total Cost 6th Iteration

Ward	d C1	d C2	Proximity
Ancol	0.2787	0.2931	0.2787
Angke	0.0215	0.0359	0.0215
Bale Kambang	0.0165	0.0304	0.0165
Bali Mester	0.0279	0.0138	0.0138
Bambu Apus	0.0157	0.0301	0.0157
Bangka	0.0037	0.0152	0.0037
Baru	0.0166	0.0295	0.0166
Batu Ampar	0.0630	0.0775	0.0630
Bendungan Hilir	0.0069	0.0195	0.0069
Bidara Cina	0.0366	0.0511	0.0366
………	…	…	…
Wijaya Kusuma	0.0533	0.0671	0.0533
Total Cost (S)			11.7978

In the 6th iteration, the value of Total Cost (S) = 11.7978
Next calculate the value of the deviation (S) with the equation
S = New Total Cost – Old Total Cost
S = 11.7978 - 7.5260
S = 4.2718

Because the value of S is greater than 0 (4.2718 > 0), the iteration calculation is stopped and the members of each cluster are taken from the 6th iteration. The selection of cluster members is done by looking at the closest distance from the centroid of the Kemayoran and Taman Sari villages. The results of manual calculations are an illustration of calculations by selecting certain data at random and cannot be compared with the results of calculations with RapidMiner. Cluster members are shown in Table 17, more details can be seen in the appendix.

Table 17. Results of Cluster 1 and 2 Members

Cluster	Village Member	Number of Members
I	Ancol	214
	Angke	
	Bale Kembang	
	Bambu Apus	
	Bangka	
	Baru	
	……	
II	Bali Mester	53
	Cideng	
	Cikini	
	Cikoko	
	Duri Selatan	
	Galur	
	……	
Total Village		267

IV. ANALYSIS

In this study, the performance of the clustering model built using the Davies-Bouldin Index. The Davies-Bouldin index aims to maximize (separate) the distance between clusters & minimize the distance between points in a cluster (dense). The value of the davies-bouldin index must be in the interval (0 and 1), the minimum value of the davies-bouldin index will indicate the optimal number of clusters. If the clustering process for each k is complete, then to determine the best number of clusters an assessment can be made using the Davies-Bouldin Index (DBI). This measurement approach aims to maximize the distance between clusters and at the same time try to minimize the distance between objects in a cluster. The cluster with the best number of clusters is the cluster with the minimum DBI value.

That is, the closer to 0 the resulting BDI value, the more optimal the resulting cluster. From the analysis that has been carried out, the DBI values in each cluster are shown in Tabel 18.

Table 18. Each DBI Value In The Simulation

Simulation to	Number of Cluster	DBI Value
1	2	0,443
2	3	0,247
3	4	0,275
4	5	0,291
5	6	0,209
6	7	0,229

Based on the results of the DBI values obtained, the cluster that has the best value is the cluster in the fifth simulation using 6 clusters with a DBI value of 0.209. In this cluster, it has the same characteristics, namely there is one cluster with a target value of vaccine administration, the number of vaccine doses 1, the number of vaccine doses 2 and the number of vaccine doses 3. These two types of clusters can be used as standard models (optimal) for grouping regions. DKI Jakarta, whose people tend to receive vaccines. A point graph with 6 clusters in the fifth simulation are shown in Figure 3 and the plot chart in Figure 4.

Figure 3. *Fifth Simulation Graphic Point*

From Figure 3 above, it can be seen that the administration of village-level vaccines for the DKI Jakarta area is still dominated by Dose 1, while Dose 3 has begun to show a level of development. From all the research that has been done by researchers, the lowest DBI value is the fifth simulation using 6 clusters with DBI. value 0.209. So if you compare the results of clustering 0, 1, 2, 3, 4 and 5 by ignoring the results of other regional clusters in DKI Jakarta, you will get the results as shown in Figure 4 below.
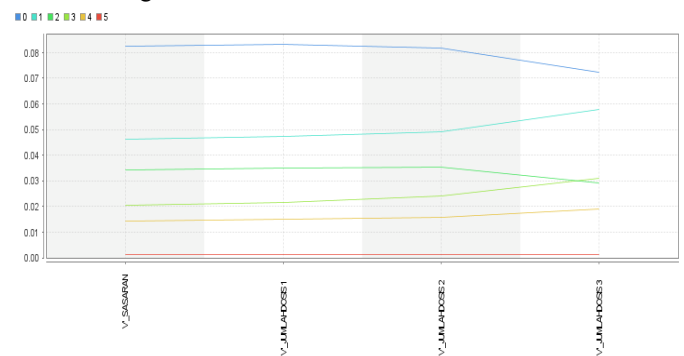

Figure 4. *Fifth Simulation Plot Chart*

In this cluster, the government still has to socialize the administration of dose 3 vaccine. In Cluster 1, the number of vaccine doses 1, 2 and 3 is considered quite successful because the graph line is above the target line. For cluster 2, especially the 3rd dose of vaccine, it still has to be socialized again even though it is close to the target target. In cluster 3, the 1st, 2nd and 3rd doses of vaccine were considered quite successful. In cluster 4, it is almost the same as in cluster 3. there is only a slight difference, namely the graph line almost does not show the distance, maybe this is influenced by the few cluster members in this cluster. In cluster 5 is the same as cluster 4. The comparison of vaccinations between clusters can be seen in Figure 5.
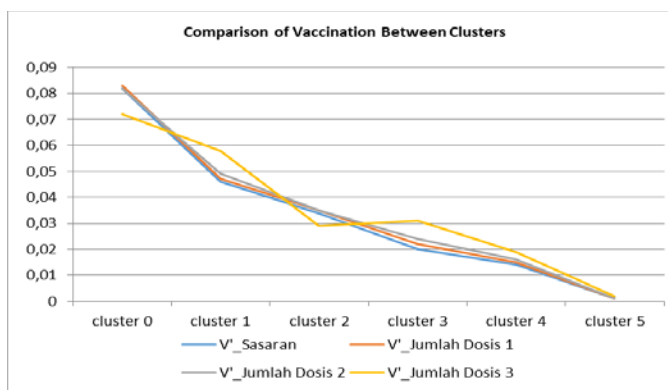
Figure 5. *Distance Comparison Graph Between Clusters*

There are two clusters that have a low vaccination rate at dose 3, namely cluster 0 and cluster 2, but cluster 2 has a lower vaccination rate than cluster 0 so that cluster 2 becomes the main priority in vaccination.

Cluster 2 is a cluster that has 67 items in the form of urban villages within DKI Jakarta. This cluster is the cluster with the highest score in vaccine dose 1, vaccine dose 2, vaccine dose 3. Number of iterations in RapidMiner is not limited, so that each variation of random data collection can produce an S value of more than 0 for each simulation carried out. Thus it can be said that all possible random data collection can be done easily.

CONCLUSION

The process of the K-Medoids Algorithm to cluster data for the DKI Jakarta City area, grouping areas for COVID-19 vaccine recipients, which was taken through a web portal provided by the DKI Jakarta Government. The data has the attributes of the DKI Jakarta urban village, the number of targets, the number of doses 1, the number of doses 2 and the number of doses 3. The processed data must be pre-processed, normalized, calculating the distance value with the Euclidean Distance equation and calculating the deviation value if the S value > 0 the cluster process is stopped. To determine the optimal number of clusters, it is seen from the smallest Davies Bouldin Index (DBI). After the calculations were done, it was found that the fifth simulation with the number of clusters 6 was the most optimal with the smallest DBI value of 0.209. The cluster chosen in this study is Cluster 2 because its members are areas in DKI Jakarta and provide a high value in COVID-19 cases compared to other clusters. This cluster has 67 items with the highest scores in all attributes. So, it can be concluded that cluster 2 is an area with vaccine doses 1, 2 and 3 doses that still need to be improved.

ACKNOWLEDGMENT

REFERENCES

[1] M. Ciotti, M. Ciccozzi, A. Terrinoni, W. C. Jiang, C. Bin Wang, and S. Bernardini, "The COVID-19 pandemic," *Crit. Rev. Clin. Lab. Sci.*, vol. 57, no. 6, pp. 365–388, 2020, doi: 10.1080/10408363.2020.1783198.

[2] M. Triana *et al.*, "Kebijakan Pemerintah Dki Jakarta Menangani Pandemi Covid-19," *J. Ilmu Pemerintah.*, vol. 14, no. 1, pp. 1–9, 2021.

[3] Noviyanto, "Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian," *J. Inform. dan Komput.*, vol. 22, no. 2, pp. 183–188, 2020.

[4] S. Sundari, I. Sudahri Damanik, A. Perdana Windarto, H. Satria Tambunan, and A. Wanto, "Prosiding Seminar Nasional Riset Information Science (SENARIS) K-Medoids Clustering Dalam Pengelompokkan Data Imunisasi Campak Balita (Siti Sundari) Analisis K-Medoids Clustering Dalam Pengelompokkan Data Imunisasi Campak Balita di Indonesia," no. September, pp. 687–696, 2019.

[5] S. Ningsih, S. Suhada, R. Dewi, and A. P. Windarto, "Pengklasteran Dokumen Tentang Dispensasi Nikah Menggunakan Algoritma K-Medoids," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 677, 2019, doi: 10.30645/senaris.v1i0.74.

[6] J. Arianto, "Penerapan Data Mining Untuk Pengelompokan Penduduk Kurang Mampu Desa Sambirejo Timur Dengan Algoritma K-Medoids (Studi Kasus Kantor Kepala Desa Sambirejo Timur)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 569–573, 2019, doi: 10.30865/komik.v3i1.1660.

[7] V. Kumar and N. Rathee, "Knowledge discovery from database using an integration of clustering and classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 2, no. 3, pp. 29–33, 2011, doi: 10.14569/ijacsa.2011.020306.

[8] F. Farahdinna, I. Nurdiansyah, A. Suryani, and A. Wibowo, "Perbandingan Algoritma K-Means Dan K-Medoids Dalam Klasterisasi Produk Asuransi Perusahaan Nasional," *J. Ilm. FIFO*, vol. 11, no. 2, p. 208, 2019, doi: 10.22441/fifo.2019.v11i2.010.

[9] H. S. Park and C. H. Jun, "A simple and fast algorithm for K-medoids clustering," *Expert Syst. Appl.*, vol. 36, no. 2 PART 2, pp. 3336–3341, 2009, doi: 10.1016/j.eswa.2008.01.039.

[10] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokkan Wilayah Sebaran Cacat pada Anak," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 64, 2018, doi: 10.24014/coreit.v4i2.4498.

[11] S. R. Joseph, H. Hlomani, and K. Letsholo, "Data Mining Algorithms: An Overview," *Int. J. Comput. Technol.*, vol. 15, no. 6, pp. 6806–6813, 2016, doi: 10.24297/ijct.v15i6.1615.

[12] E. Schubert and P. J. Rousseeuw, "Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA, and CLARANS algorithms," *Inf. Syst.*, vol. 101, p. 101804, 2021, doi: 10.1016/j.is.2021.101804.

[13] F. Virgantari and Y. E. Faridhan, "in Indonesia ' s Provinces," vol. 5, no. 2, pp. 1–7, 2020.

[14] Syahrul dkk, "Teknologi informasi dan pendidikan," *Al-Manar (Edisi 1)*, vol. 12, no. 2, pp. 1–7, 2004.

[15] S. Haryati, A. Sudarsono, and E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.

[16] R. Adhitama, A. Burhanuddin, and R. Ananda, "Penentuan Jumlah Cluster Ideal Smk Di Jawa Tengah Dengan Metode X-Means Clustering Dan K-Means Clusterin," *JIKO (Jurnal Inform. dan Komputer)*, vol. 3, no. 1, pp. 1–5, 2020, doi: 10.33387/jiko.v3i1.1635.

[17] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *J. Sains dan Manaj.*, vol. 9, no. 1, pp. 95–100, 2021, [Online]. Available: https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/10428/4839.

[18] M. A. Fajar, "Klasterisasi Proses Seleksi Pemain Menggunakan Algoritma K-Means," *J. Tek. Inform.*, pp.

1–5, 2015, [Online]. Available: http://eprints.dinus.ac.id/16498/1/jurnal_15441.pdf (05 Juni 2020).

[19] K. Aulia, R. Hartanto, and S. Fauziati, "Analisis Faktor-faktor yang Mempengaruhi Pengelolaan Website pada Pemerintah Daerah Istimewa Yogyakarta," *Semin. Nas. Ilmu Komput.*, no. Snik, pp. 1–8, 2016.

[20] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. Ilmi R.H.Zer, and D. Hartama, "Analisis algoritma K-Medoids clustering dalam pengelompokan penyebaran Covid-19 di Indonesia," *Jti (Jurnal Teknol. Informasi)*, vol. 4, no. 1, pp.