ISSN : 1978-8282, Online ISSN: 2655-4275

Comparison of the K-Means Method with the K-Medoids Method for Recipient Infant Clustering Measles Immunization

Muhammad Faisal[*1], Wiranti Sri Utami[2], Vera Nurlia[3], Nila Pratiwi[4]
[1] Informatics Engineering Study Program, Faculty of Science & Technology, University of Raharja, Indonesia
[2,3] Informatics Engineering Study Program, Faculty of Engineering, Cendekia Abditama University, Indonesia
[4] Electrical Engineering Study Program, Faculty of Engineering, Cendekia Abditama University, Indonesia
E-mail: [*1]**muhammad.faisal@raharja.info**, [2]wirantisutami@uca.ac.id, [3]veranurlia@uca.ac.id, [4]nilapratiwi@uca.ac.id

Abstract

Measles or meales is a disease that can be transmitted, if someone who is affected by measles is in the same room with other people who have never been exposed to or received the measles vaccine, it is estimated that they will be infected with measles. The measles vaccine is a very effective vaccine for preventing measles. This type of vaccine is included in the routine immunization program recommended by the Ministry of Health of the Republic of Indonesia. So far, there are two types of measles vaccine, namely the MR vaccine and the MMR vaccine. The MR vaccine is used to prevent measles and rubella, while the MMR vaccine prevents measles, rubella and mumps. The target of vaccine administration can be observed by applying the data mining method, while the data mining method that can be used in this study is by using K-Means and K-Medoids. The purpose of this study was to compare the K-Means and K-Medoids methods in analyzing the achievement of vaccine administration to infants aged 9 months to 18 months. The results of the study obtained optimal k values of 21 provinces with the highest cluster of 3 using the K-Means algorithm, while the optimal K values of 10 provinces with the highest cluster of 3 were obtained using the K-Medoids algorithm.

Keywords *— Measles Immunization, Clusterring, K-Means, K-Medoids*

1. INTRODUCTION

Measles is an infectious disease caused by the measles virus, this type of disease is a health problem that must be taken seriously because the number of cases is quite high and is included in the category of Extraordinary Events[1]. Therefore a special treatment is needed to prevent the measles virus, immunization programs are an important solution to prevent measles. Measles itself is included in the PD3I category (Diseases That Can Be Prevented By Immunization)[2]. According to the WHO, measles is an infectious disease with symptoms of redness in the form of spots for 3 hearts and a fever of 380C[3].

To overcome the spread of the measles virus, the Government has carried out a vaccine or immunization program, the measles vaccine is considered effective in preventing measles. This research will focus on administering measles immunization to infants aged twelve months

and under. Infants aged less than one year are more susceptible to measles than adults. This grouping process will be divided into three categories where each year the administration of measles immunization will be carried out by research using data mining.

Data mining is a separation technique between useful information models stored in databases, data mining is also a data processing technique that functions to find information that is not yet in the database[4]. In the data mining process there are several methods for data processing techniques, the method that will be used in this research is to use K-Means which will be compared with K-Medoids.

2. RESEARCH METHOD

The methods to be used in this study include the following.

2.1. Data Mining

Data mining is a process of artificial intelligence, statistical techniques, mathematics, and machine learning that is used to identify and break down a data that has useful information contained in the database. Data mining is included in the KDD (Knowledge Discovery in Database) category of techniques, KDD is a method used to obtain knowledge originating from databases [5].

2.2. Clustering

Clustering is a method for grouping data that has similarities into certain groups [6]. In using the clustering method there are several algorithms including K-Means and K-Medoids which will be discussed in this study.

2.3. K-Means

K-means is one of the algorithms contained in the clustering technique, K-Means is a simple method of conducting grouping analysis which aims to determine data grouping and separate an entity into a group called a cluster [7].

2.4. K-Medoids

The K-Medoids algorithm is a clustering technique that is used to collect a set of objects into a set of clusters, K-Medoids is a Patition Around Medoids (PAM) algorithm developed by Leonard Kaufman and Peter J. Rousseeuw in 1987 [8]. Medoids are a PAM cluster representation of a set of objects representing a cluster.

3. RESEARCH RESULTS AND DISCUSSION

The data to be used in this study uses data from the Central Bureau of Statistics (BPS). This data is the percentage of children under five who have received measles immunization in each province in the Republic of Indonesia.

Table 1. Percentage of Toddlers Receiving Measles Immunization

Province	Percentage of Toddlers Who Have Received Measles Immunization (Percent)		
	2020	2021	2022
Aceh	39.21	38.11	38.19
Sumatera Utara	62.72	62.32	65.07
Sumatera Barat	56.40	56.03	56.48
Riau	60.96	63.38	62.31
Jambi	63.47	65.63	63.65
Sumatera Selatan	69.00	70.52	74.00
Bengkulu	75.38	76.94	76.41
Lampung	73.18	75.83	76.18
Kep. Bangka Belitung	65.47	71.52	70.36
Kep. Riau	68.47	71.78	75.32
DKI Jakarta	72.82	72.31	74.19
Jawa Barat	67.47	66.23	69.03
Jawa Tengah	73.36	75.30	75.67
DI Yogyakarta	78.67	77.50	77.88
Jawa Timur	70.67	70.18	71.25
Banten	58.65	63.81	65.55
Bali	80.18	78.37	81.69
Nusa Tenggara Barat	73.07	74.07	76.29
Nusa Tenggara Timur	74.39	76.46	78.75
Kalimantan Barat	65.88	65.02	64.65
Kalimantan Tengah	65.52	66.58	67.72
Kalimantan Selatan	67.89	68.63	67.75
Kalimantan Timur	68.86	73.01	70.55
Kalimantan Utara	71.63	72.63	72.67
Sulawesi Utara	74.39	75.39	77.05
Sulawesi Tengah	69.85	71.21	71.60
Sulawesi Selatan	73.39	73.21	73.84
Sulawesi Tenggara	69.82	73.78	74.81
Gorontalo	75.13	76.43	72.99
Sulawesi Barat	66.76	68.23	67.92
Maluku	68.78	69.54	72.76
Maluku Utara	65.74	69.22	69.45
Papua Barat	69.08	67.95	70.99
Papua	51.84	60.75	69.25

3.1. K-Means Method

In implementing the K-Means method, the author will represent the K-Means method using R-Studio Cloud, which has now changed its name to Posit Cloud. The initial stage before using this application is to install several packages to support the cluster search process and find the optimal K value then execute K-Means.

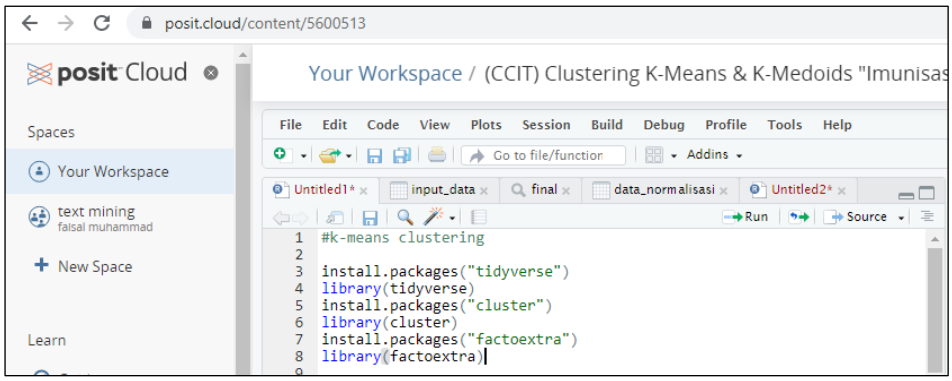


Figure 1. Install Packages Needed for K-Means Clusters

The next stage is preparing the dataset to be analyzed using Posit Cloud, these steps are uploading data to Posit Cloud, cleaning the data, and selecting data to be processed by clustering using Posit Cloud.

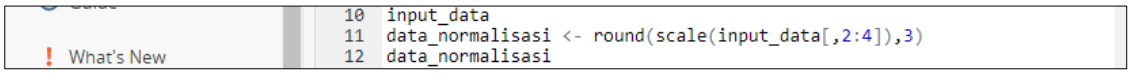


Figure 2. Data Upload and Data Normalization Process

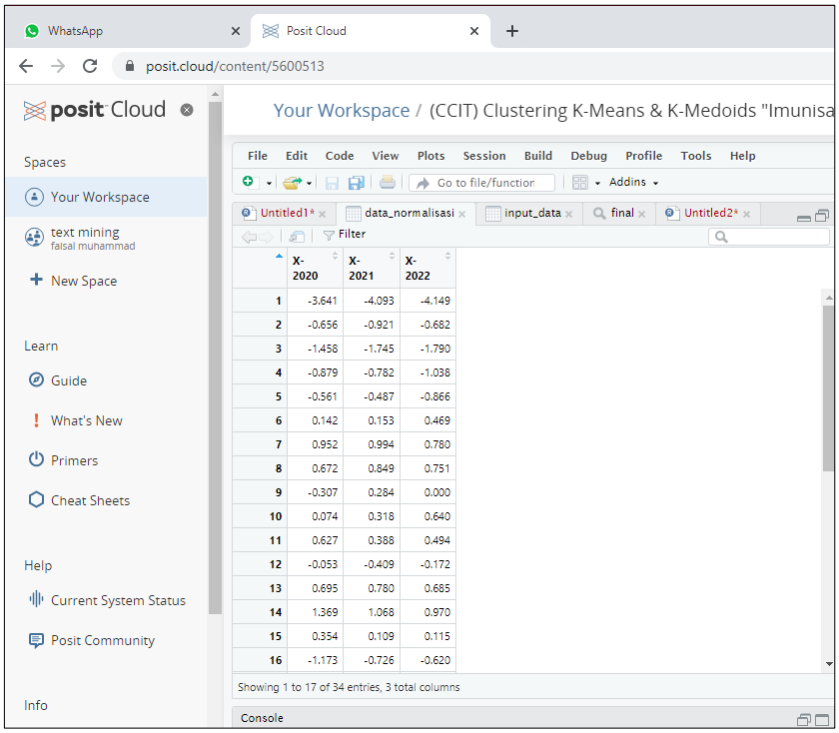


Figure 3. K-Means Data Normalization

The next stage is to find the optimal K value, the normalized data will enter the preparation process using the elbow method in the Posit Cloud application.

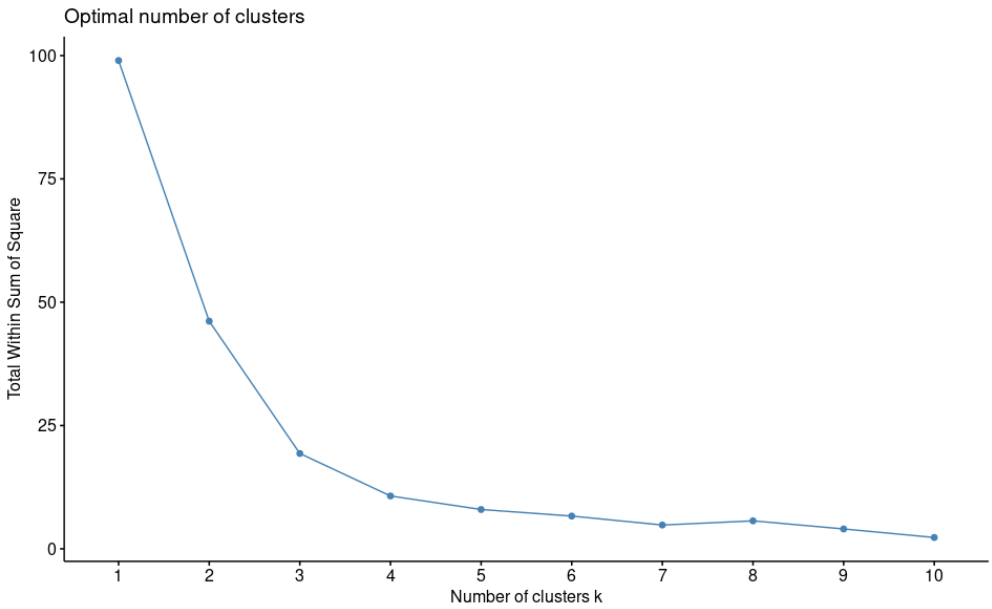


Figure 4. Graph of the Results of the K-Means Cluster Optimal K Value

The picture above shows that the optimal K value occurs at a decrease of 3, so in this study using K-Means using K 3 or 3 clusters which will be used in this study using Posit Cloud.

The next step is to determine the algorithm calculation using K-Means, because the K-Means cluster has been determined, the K-Means algorithm calculation process can be carried out on the Posit Cloud application.

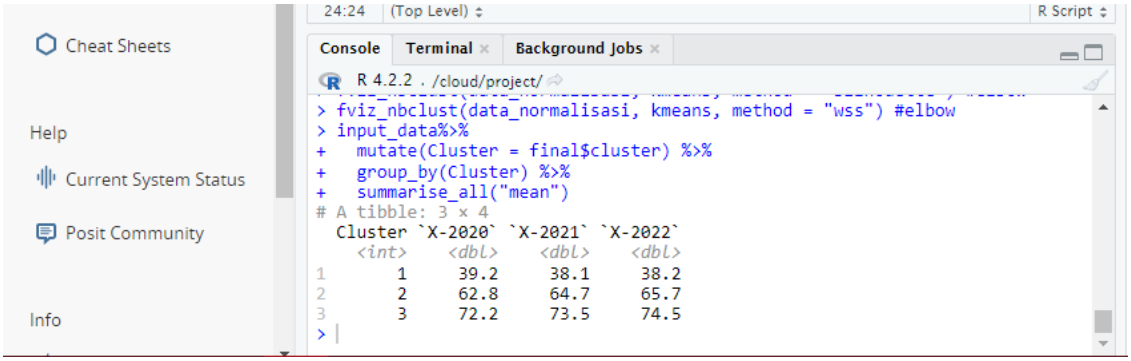


Figure 5. The Final Result of the K-Means Cluster

3.2. K-Medoids Method

Then in the process of using the K-Medoids method the first step is to install the package that will be used to search for K values using the Posit Cloud application. Then upload data, clean data, and normalize data, the process can be seen in the image below.

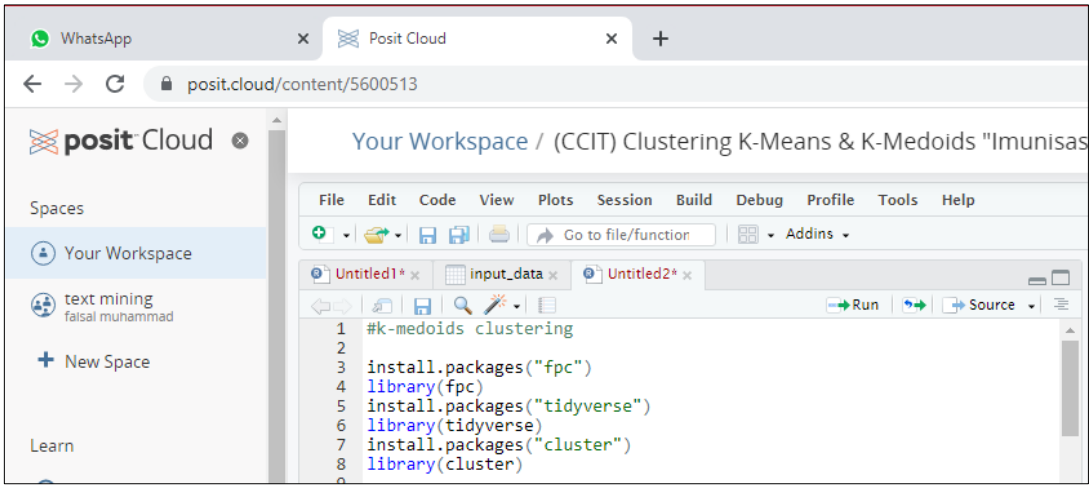
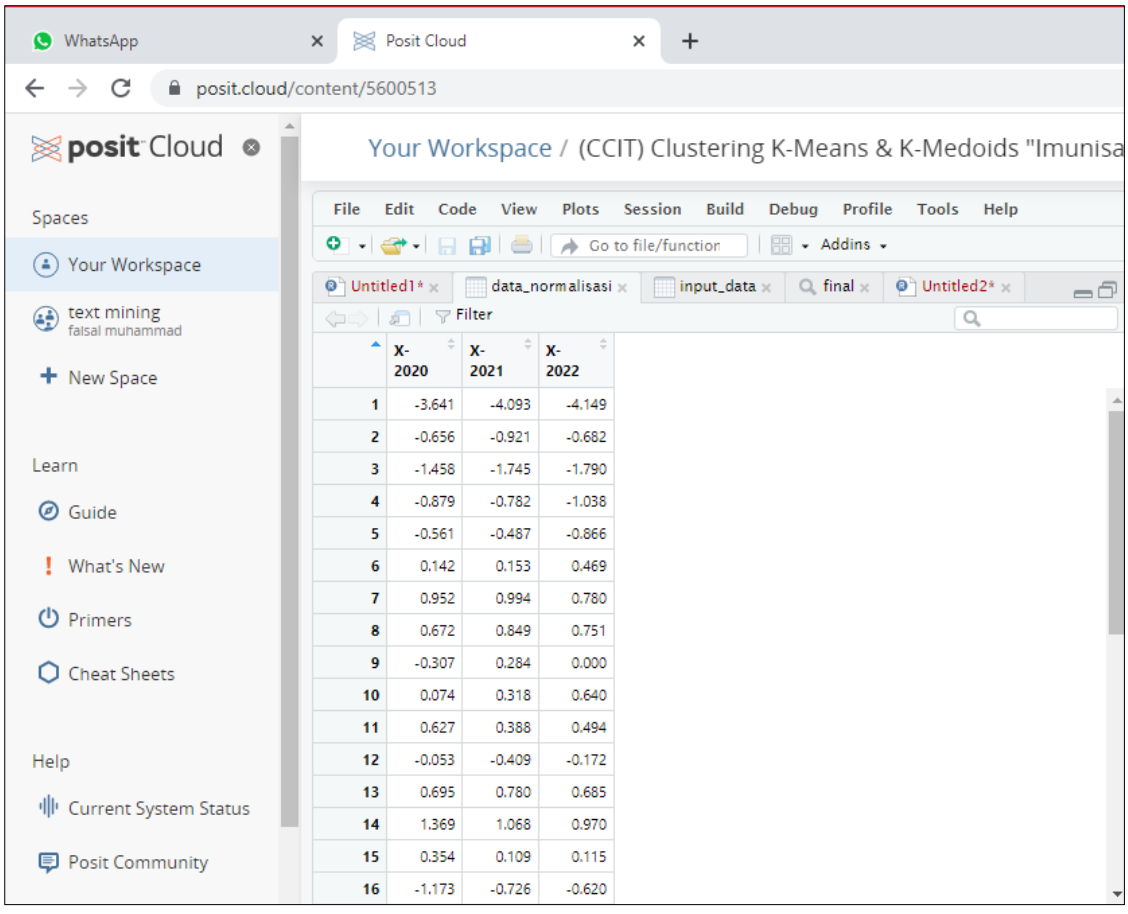


Figure 6. Install the Packages Needed for the K-Medoids Cluster

The next step is to normalize the dataset against the data that has been provided using the K-Medoids cluster, the process can be seen in the image below.



	X-2020	X-2021	X-2022
1	-3.641	-4.093	-4.149
2	-0.656	-0.921	-0.682
3	-1.458	-1.745	-1.790
4	-0.879	-0.782	-1.038
5	-0.561	-0.487	-0.866
6	0.142	0.153	0.469
7	0.952	0.994	0.780
8	0.672	0.849	0.751
9	-0.307	0.284	0.000
10	0.074	0.318	0.640
11	0.627	0.388	0.494
12	-0.053	-0.409	-0.172
13	0.695	0.780	0.685
14	1.369	1.068	0.970
15	0.354	0.109	0.115
16	-1.173	-0.726	-0.620

Figure 7. K-Medoids Data Normalization

The next step is to find the K Optimal cluster value to be used in the clustering process using K-Medoids. This process can be seen in the image below.

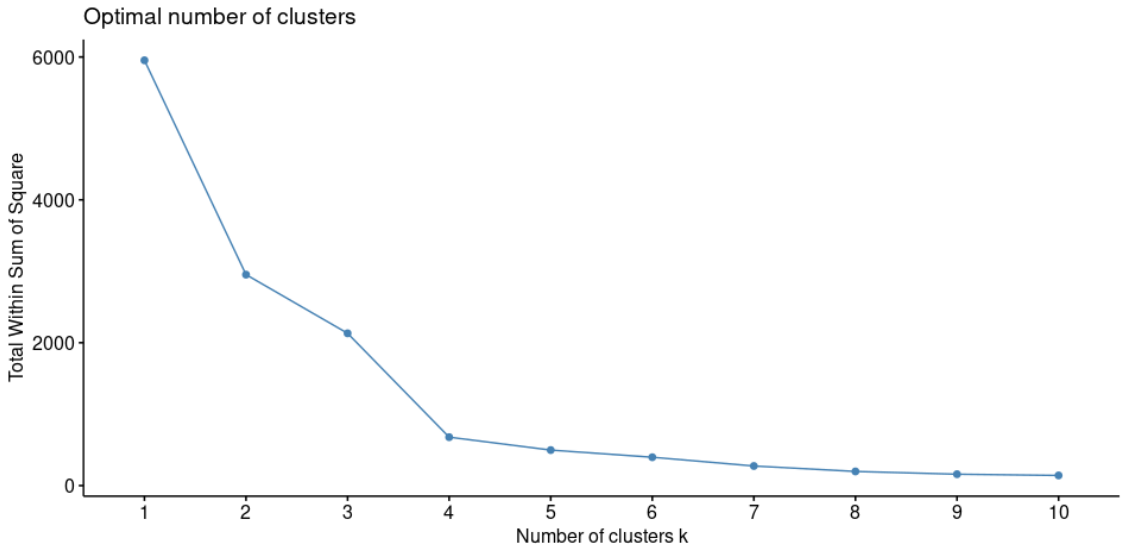


Figure 8. Graph of the Results of the K-Medoids Cluster Optimal K Value

In the process of finding the optimal K value from the graph above, it shows that the decrease in K values occurs in K 3 values or cluster 3. The next step is to find the final value of the K-Medoids cluster. Because the K value has been determined, the search process for K values can be carried out in the Posit Cloud application.

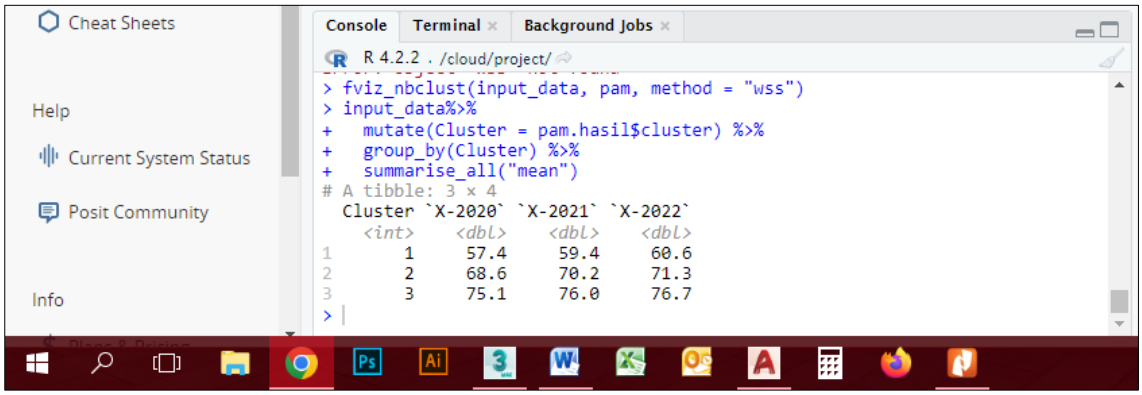


Figure 9. The Final Result of the K-Medoids Cluster

3.3. Analysis Results

The results of the comparative analysis process for Optimal K values using the K-Means and K-Medoids algorithms, it was found that the highest K Optimal values in the K-Means algorithm were obtained in 21 Provinces with K Optimal clusters of 3, while the Optimal K values in the K-Algorithm The highest medoids were obtained in 10 provinces with a cluster value of 10.

Table 2. Display of Optimal K for each cluster of Provinces Recipient of Measles Immunizationin 2020, 2021, 2022 using the K-Means and K-Medoids Algorithm

Province	K-Means	K-Medoids
Aceh	2	1
Sumatera Utara	1	1
Sumatera Barat	1	1
Riau	1	1
Jambi	1	1
Sumatera Selatan	3	2
Bengkulu	3	3
Lampung	3	3
Kep. Bangka Belitung	3	2
Kep. Riau	3	2
DKI Jakarta	3	2
Jawa Barat	1	2
Jawa Tengah	3	3
DI Yogyakarta	3	3
Jawa Timur	3	2
Banten	1	1
Bali	3	3
Nusa Tenggara Barat	3	3
Nusa Tenggara Timur	3	3
Kalimantan Barat	1	1
Kalimantan Tengah	1	2
Kalimantan Selatan	1	2
Kalimantan Timur	3	2
Kalimantan Utara	3	2
Sulawesi Utara	3	3
Sulawesi Tengah	3	2
Sulawesi Selatan	3	3
Sulawesi Tenggara	3	2
Gorontalo	3	3
Sulawesi Barat	1	2
Maluku	3	2
Maluku Utara	1	2
Papua Barat	3	2
Papua	1	1

4. CONCLUSION

The results of this study K Optimal from the Algorithm of K-Means values are in 21 Provinces with Optimal values of 3 Clusters including South Sumatra, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Central Java, DI. Yogyakarta, East Java, Bali, NTB, NTT, East Kalimantan, North Kalimantan, North Sulawesi, Central Sulawesi, Southeast

Sulawesi, Gorontalo, Maluku, West Papua. While the Optimal K value in the K-Medoids algorithm is in 10 Provinces with Optimal values of 3 Clusters including Bengkulu, Lampung, Central Java, DI. Yogyakarta, Bali, NTB, NTT, North Sulawesi, South Sulawesi and Gorontalo.

5. SUGGESTED

The author hopes that there will be further research related to the case study that the author uses using other analytical techniques such as Data Warehouses or Decision Support Systems.

6. REFERENCES

[1] P. N. Yahmal, "FAKTOR-FAKTOR YANG BERHUBUNGAN DENGAN KEJADIAN CAMPAK," Jurnal Medika Hutama , vol. 03, no. 01, pp. 1612-1615, 2021.

[2] Y. N. M. Sevilla Ukhtil Huvaid, "Faktor-faktor yang Berhubungan dengan Pemberian Imunisasi," Jurnal Riset Hesti Medan Akper Kesdam I/BB Medan, vol. 4, no. 2, pp. 82-87, 2019.

[3] T. Humas, "https://www.rsuppersahabatan.co.id/," Rumah Sakit Umum Persahabatan, 14 Oktober 2019. [Online]. Available: https://www.rsuppersahabatan.co.id/artikel/read/campak#:~:text=Pengertian%20campak%20menurut%20WHO%20adalah,batuk%2C%20(Cough)%2C%20bercak. [Accessed 16 03 2023].

[4] R. R. K. K. lra Zulfa, "Implementasi data mining untuk menentukan strategi penjualan buku bekas dengan pola pembelian konsumen menggunakan metode Apriori (studi kasus: Kota Medan)," JURNAL SAINS DAN TEKNOLOGI , vol. 16, no. 1, pp. 69-82, 2020.

[5] M. A. Y. D. Gustientiedina, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," Jurnal Nasional Teknologi dan Sistem Informasi, vol. 05, no. 1, pp. 17-24, 2019.

[6] S. Nur Aini Fadhila, "Perbandingan Algoritma K-Means dan K-Median Clustering Terhadap Nilai Ujian Nasional SMP di Jawa Tengah," Jurnal Smart Comp, vol. 11, no. 4, pp. 563-572, 2022.

[7] Y. N. Erdi Amos Saputra, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," Journal of Information Systems and Informatics , vol. 3, no. 3, pp. 424-439, 2021.

[8] E. H. A. N. S. R. Evi Rahmah, "Penerapan Algoritma K-Medoids Clustering Untuk Menentukan Srategi Promosi Pada Data Mahasiswa (Studi Kasus : Stikes Perintis Padang)," Jurnal Nasional Komputasi dan Teknologi Informasi, vol. 5, no. 3, pp. 556-564, 2022.