Journal of Information Systems and Informatics
Vol. 4, No. 3, September 2022 e-ISSN: 2656-4882 p-ISSN: 2656-5935
http://journal-isi.org/index.php/isi Published By DRPM-UBD

Population Density Cluster Analysis in DKI Jakarta Province Using K-Means Algorithm

Amalia Anjani Arifiyanti[1], Farhan Setiyo Darusman[2], Brahmantio Widyo Trenggono[3]

[1,2,3] Program Studi Sistem Informasi, Universitas Pembangunan "Veteran" Jawa Timur, Surabaya, Indonesia
Email: [1]amalia_anjani.fik@upnjatim.ac.id, [2]setiyofarhan@gmail.com, [3]brahmantio25@gmail.com

Abstract

This study aims to analyze clusters based on the area and population density of the area and population density of the area in DKI Jakarta Province in 2015 using the data mining method by clustering as the first step in planning for population equality. The subject of analysis in this study is a village located in the province of DKI Jakarta which is recorded based on the area and population density in each sub-district until 2015 with several stages, namely data understanding, data processing or cleansing, cluster tendency assessment, clustering, cluster review. From this study, the results were obtained that the data tended to be clustered because the statistical value of Hopkins was close to the value of 0 and in VAT there was a vague picture of clusters that might be formed. Based on this, cluster creation is carried out using the K-Means Algorithm. Based on the results, there are 3 clusters formed, namely cluster 0 (not densely populated), cluster 1 (medium population density), and cluster 2 (densely populated). These results can be used as a basis for policy making in population management.

Keywords: Clustering, Data Mining, Population Density, K-Means

1. INTRODUCTION

The area of DKI Jakarta is around 662.33 km² with a population density of DKI Jakarta in 2020 reaching 16,704 people / km²[1], [2] . of course, this makes the DKI Jakarta area filled with many residents rather than the predetermined area. In addition, based on existing regulations and regional boundaries, DKI Jakarta is divided into 5 Municipality areas and 1 Administrative Regency, namely Central Jakarta, West Jakarta, East Jakarta, South Jakarta, North Jakarta, and the Thousand Islands. The DKI Jakarta area itself is in the south of the Java Sea, then the east side which borders Bekasi city, then if the south is bordered by Bogor City, then the west is bordered by Tangerang City. With each of the 6 regions it is headed by an elected Mayor. The total area of DKI Jakarta itself has 43 sub-districts and 268 sub-districts [3], [4].

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Data collection every year will continue to be carried out by the competent authorities. The data collection is carried out to find out the growth of the community, especially in DKI Jakarta which is getting denser from year to year. Data collection is also carried out based on the villages in each region. From these data, it can be known which areas have an increasing population every year. As stipulated in Law Number 24 of 2013 concerning Population Administration, population data collection is carried out for public services, development planning, allocation, democratic development, law enforcement and crime prevention. [5]

The problem of overcrowding occurs in several urban villages in DKI Jakarta. However, in several other villages, the population density is still low compared to the area of the village. This data was obtained from population data from the DKI Jakarta Provincial Population and Civil Registration Service which was made on April 25, 2016. An example of the uneven population density in DKI Jakarta Province is as happened in Jelambar Baru and South Grogol villages which experienced a population density with a population density of more than 150 thousand but an area of less than 1 km2. On the other hand, there are villages such as Rorotan and Kamal Muara whose area is more than 10 km2 but the population is less than 5000. If the population density is classified into three classifications of High, Medium, and Low based on the average level of all population density in DKI Jakarta Province, the result is that Jelambar Baru village is included in the High category, South Grogol is included in the medium category, and other areas are included in the Low category. This happens because of the difference in density levels that are very far in the two regions compared to other regions in DKI Jakarta Province. It also shows that the distribution of the population is not evenly distributed. This is quite different when viewed based on the category of population density according to the Ministry of PUPR. 42% of DKI Jakarta Province is included in the High-density category, 19% is included in the medium density category, and the remaining 40% is included in the Low-density category. A fragment of the density status from the DKI Jakarta area data can be seen in table 1 below with R being Low density, S being medium density, and T being High density.[6][7]

Table 1. Examples of population density data

No	Regency	District	Neighborhoods	Area (km^2)	Density ($soul/km^2$)	Density Status
1	East Jakarta	Makasar	Halim perdana kusumah	13.07	2623.41	R
2	Central Jakarta	Gambir	Cideng	1.26	15108.40	S
3	West Jakarta	Palmerah	Palmerah	2.33	30680.25	T
4	West Jakarta	Tambora	Pekojan	0.78	35262.21	T

5	Central Jakarta	New Johar	High ground	0.62	73374.54	T

These various strategies and planning for population equality cannot be done directly without a basis of data reference. One of the initial stages is to understand the current conditions, namely by analyzing data on population, especially regarding data on population density levels in each region. One way that can be used to support the analysis of data conditions is by dividing or grouping regions and in this study is based on the level of population density. This is done so that it can be known how many regional groups and which regions are included in the group so that they can be used as recommendations and considerations in the preparation of the next strategy. One of the methods in the division or grouping of data to describe data is with [8], [9]clustering analysis techniques.[10]

Population density is something that needs to be analyzed far so that it can be managed so that it does not cause further problems. Population data collection is carried out automatically through information systems starting from can be one of the methods to support monitoring the movement of population density. Further population density analysis can be carried out by the cluster analysis method. This cluster analysis is carried out by the method of clustering. This method is stated to have succeeded in forming population density clusters in various regions in Indonesia with diverse attribute approaches, such as ranging from population attributes to gender attributes, age, human development index ([18][19]–[21][22][23]Human Development Index), school participation rate, and unemployment rate. With most of the clustering is done with the K-Means algorithm.[24]

The K-Means algorithm can be used to form a cluster if the cluster that is expected to form is not hierarchical. In addition, there are several other provisions in the application of this algorithm, namely the number of clusters that are expected to be formed has been determined in advance and the data attributes used must be numerical or numerical. For the hatching of the number of clusters can be done in several ways for example the elbow method, silhouette, and so on. Cluster formation with the K-Means algorithm uses the distance calculation method. Distance calculations generally use [11][25][26]the Euclidean method (Euclidean Distance), although nevertheless this distance calculation method can be modified using other distance calculation methods[25][27].

Identification of urban clusters based on area and population density in DKI Jakarta province in this study was carried out with the K-Means Algorithm. The formation of regional clusters with K-Means was chosen because the clusters formed are partial or non-hierarchically suitable for this study. In general, cluster formation is carried out by grouping a set of objects that have a high level of

similarity compared to other objects in different groups or clusters. The calculation of the degree of similarity in the K-Means algorithm is to calculate the distance between objects. The closer the distance between objects, the objects are considered to have a high level of similarity, on the other hand, if the distance between objects is far away, it is considered that the objects have a low level of similarity[11][12].

2. METHOD

This research is a data mining method with a focus on cluster analysis. The stages of this study can be seen in figure 1. This research began with determining the purpose of the study, namely the analysis of the DKI Jakarta area cluster based on its population density. Based on this goal, the next stage is the search and collection of data on population density in DKI Jakarta. After the data is collected, the data needs to be understood by paying attention to what aspects of data attributes are available and what are needed and the quality of the data. At the data processing stage, data cleaning, duplication deletion, and other activities were carried out to improve data quality and the selection of attributes to be used in this study was carried out. The next stage is a cluster tendency assessment or cluster tendency assessment which is carried out to see if there is a cluster tendency in the data that is the subject of the study. If it is judged that there is a tendency to clusters, the next step is to determine the number of clusters to be formed by the elbow and silhouette methods to find the most optimal number of clusters. Furthermore, cluster formation is carried out where the parameters for this clustering are based on the number of inhabitants and area according to the attributes in the dataset, which is then followed by the analysis stage of the cluster formed.

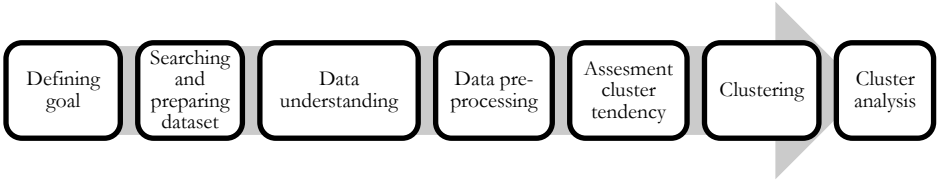


Figure 1. Research Framework

At the stage of cluster formation, the K-Means Clustering algorithm is used. K-Means Clustering is one of the algorithms that is widely used in carrying out cluster formation. The stages of K-Means algorithm in the Clustering process are depicted in Figure 2. The distance calculation in this study uses the Euclidean distance formula [28].

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

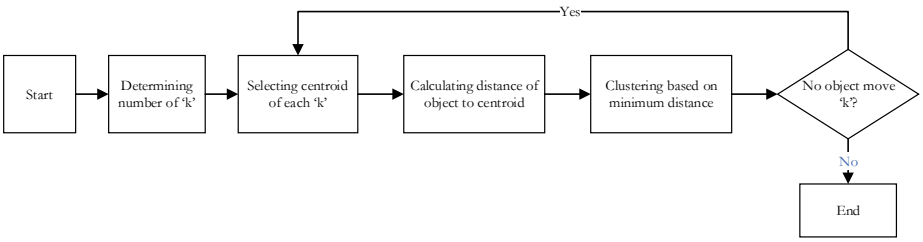


Figure 2. Workflow K-Means Clustering Algorithm

3. RESULT AND DISCUSSION

The following is the result of the work on this research whose stages have been described in the previous section.

3.1 Search and Data Collection

The data that will be used in this study comes from the official website of the Integrated Data Portal of the DKI Jakarta Provincial Government, namely datajakarta.go.id. This data is population data obtained from the DKI Jakarta Provincial Population and Civil Registration Service, which was made on April 25, 2016, and last updated on May 23, 2018.

3.2 Understanding Data

The data used consists of 267 rows with attributes: Year, nama_provinsi, nama_kabupaten, nama_kecamatan, nama_kelurahan, luas_wilayah_(km2), kepadatan_(soul/km2). In this study, clustering will only use two attributes, namely luas_wilayah_ (km2), kepadatan_ (soul / km2). Here is an example of the data for which the analysis will be carried out, presented in Table 2.

Table 2. Example of dataset fragments

Name regency	Name district	Name Neighborhoods	Area (km2)	Density (soul/km2)
East Jakarta	Makasar	Halim perdana kusumah	13.07	2623.41
Central Jakarta	Gambir	Cideng	1.26	15108.40
West Jakarta	Palmerah	Palmerah	2.33	30680.25
West Jakarta	Tambora	Pekojan	0.78	35262.21
Central Jakarta	New Johar	High ground	0.62	73374.54

The description of the data can be seen in figure 3 (a). From the description of the data, it was obtained that there are 267 rows and 2 columns that will be used later.

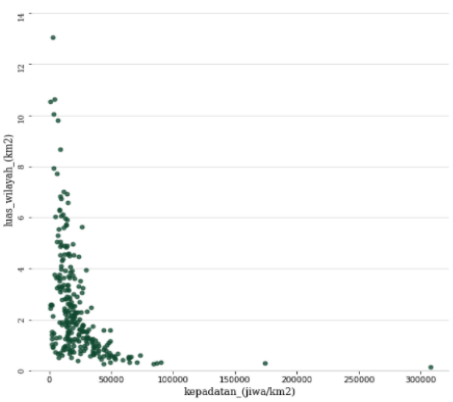
p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

The data type on both columns to be used is float. In this dataset, there is also no blank data on both attributes.

Furthermore, outlier detection of the data is carried out using scatterplot. The results of the scatterplot chart from the data used in this study can be seen in figure 3 (b) below. From the scatter chart, there are some outlier data that can affect the clustering results, namely data that has a population density value of >80,000. This outlier data will be processed at a later stage.

```
   kepadatan_(jiwa/km2)  luas_wilayah_(km2)
0              10278.58                0.62
1               2539.95                2.58
2                962.32                2.45
-----------------
Shape:  (267, 2)
-----------------
kepadatan_(jiwa/km2)    float64
luas_wilayah_(km2)      float64
dtype: object
-----------------
       kepadatan_(jiwa/km2)  luas_wilayah_(km2)
count            267.000000          267.000000
mean           23946.115393            2.449288
std            25492.494027            2.073503
min              962.320000            0.140000
25%            11913.830000            0.965000
50%            17579.120000            1.790000
75%            29751.295000            3.300000
max           307642.560000           13.070000
```

(a)



(b)

Figure 3. Dataset; (a) Description of the data; (b) Garfik scatter

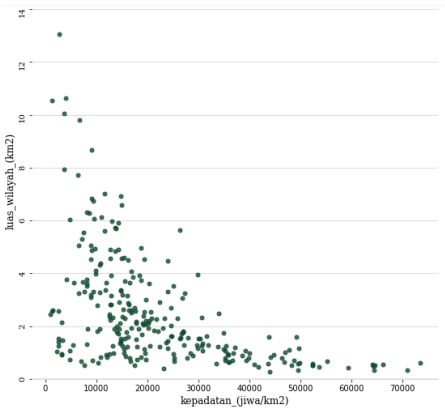
3.3 Data Processing

In the previous stage, the outlier data in the dataset was identified. This outlier data was deleted so that there were some changes, namely the number of rows of data that was previously 267 reduced to 262, and the overall statistical value. The significant change that occurred was the max value, which was previously above 300,000, currently changing to 73374.54. The overall results of the data description after the outlier deletion process can be seen in figure 4 (a). From the scatterplot chart (Figure 4 (b)) after the outlier data is removed, the scatter plot chart looks clearer, and the data is assessed as ready for clustering.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

```
   kepadatan_(jiwa/km2)  luas_wilayah_(km2)
0              10278.58                0.62
1               2539.95                2.58
2                962.32                2.45
-----------------
Shape:  (262, 2)
-----------------
kepadatan_(jiwa/km2)    float64
luas_wilayah_(km2)      float64
dtype: object
-----------------
       kepadatan_(jiwa/km2)  luas_wilayah_(km2)
count            262.000000          262.000000
mean           21568.411756            2.491031
std            14701.599094            2.070821
min              962.320000            0.280000
25%            11788.425000            0.990000
50%            17286.575000            1.855000
75%            27856.197500            3.315000
max            73374.540000           13.070000
```

(a)



(b)

Figure 4. Data from outlier deletion processing; (a) Data description); (b) Scatter chart

3.4 Cluster Tendency Assessment

At this stage, a data tendency assessment of the clustering method will be carried out using a Virtual Assessment of cluster Tendency (VAT) and Hopkins statistics whose results can be seen in figure 5. Nilai Hopkins statistic obtained is 0.3585, it can be concluded that this data tends to be clustered because the value is close to 0. This is reinforced by VAT where on the VAT chart it can show a vague picture of the clustering that will be formed later.

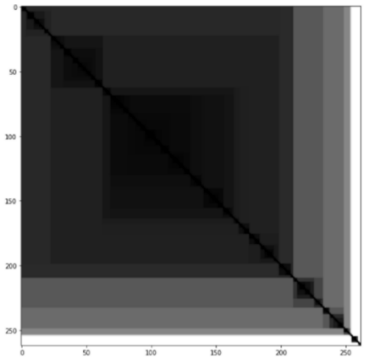


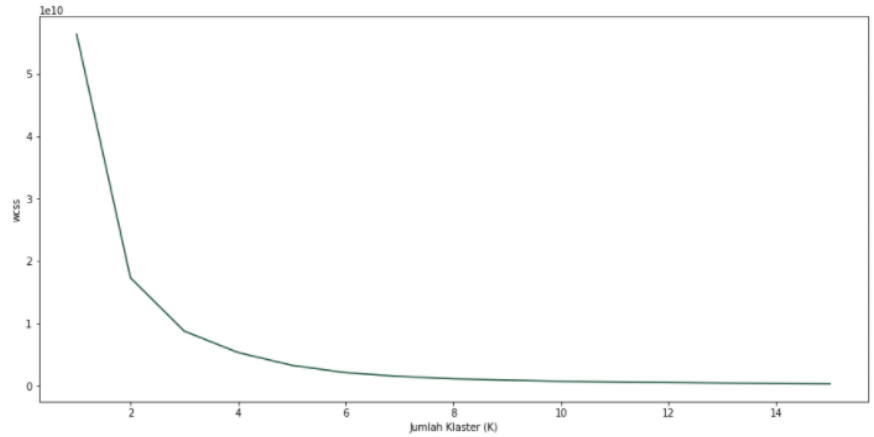
Figure 5. VAT chart

3.5 Cluster Formation

After the data is assessed, it tends to be clustering. Furthermore, the clustering process is carried out using K-Means Clustering. In conducting K-Means Clustering the first thing to do is to determine the number of clusters to be formed.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

To do so will be used elbow and silhouette methods. The results of the elbow method and the silhouette value can be seen in figure 6 [29]. From the elbow chart, the elbow point is at point 3, this means that the optimal cluster value for this data is 3. This is also supported by the silhouette method where the number of clusters 3 has the highest average silhouette value compared to the number of other clusters.




```
Jumlah Klaster = 3 nilai rata-rata silhouette = 0.5960686020171135
Jumlah Klaster = 4 nilai rata-rata silhouette = 0.5513647272523509
Jumlah Klaster = 5 nilai rata-rata silhouette = 0.5740932262450616
Jumlah Klaster = 6 nilai rata-rata silhouette = 0.5877551424217222
Jumlah Klaster = 7 nilai rata-rata silhouette = 0.5597797568213458
```

Figure 6. The results of the elbow and silhouette score method

After obtaining the number of clusters to be formed, cluster formation will be carried out using the K-Means algorithm. Cluster results based on their centroid can be seen in table 3.

Table 3. Centroid Cluster

Cluster	Density (Soul/Km²)	Area (Km²)
0	18690.1	2.8
1	23275.6	2.3
2	36089.2	1.2

3.6 Cluster Analysis

The visualization of the formed clusters can be seen on the two-dimensional Scatter Plot Chart in figure 7.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

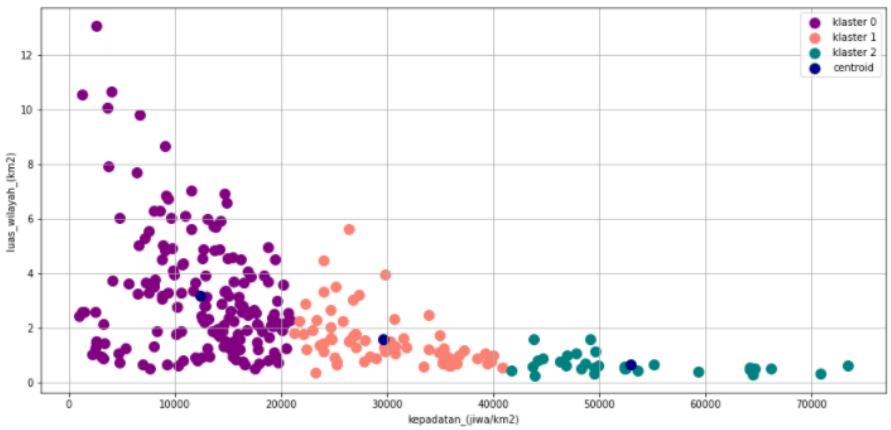


Figure 7. Cluster Visualization

From the graph in figure 7, there are three clusters formed which are divided based on the value of the density of life / km². Where the three clusters are can be articulated as follows:

Cluster 0: low density of people / km² (< 20000)
Cluster 1: density of life / km² medium (20000 – 40000)
Cluster 2: density of people / km² high (> 40000)

From these data, it can be concluded that Cluster 2 is a densely populated cluster, Cluster 1 is a medium population density cluster, and Klaster 0 is a cluster with a low population density. Table 4 contains sample data from each cluster.

Table 4. Sample Data with Cluster Results

Name regency	Name district	Name Neighborhoods	Area (km2)	Density (soul/km2)	Cluster Results	Criteria based on PUPR
East Jakarta	Makasar	Halim perdana kusumah	13.07	2623.41	0 / R	R
North Jakarta	Cilincing	Cideng	1.26	15108.40	0 / R	S
West Jakarta	Palmerah	Palmerah	2.33	30680.25	1 / S	T
West Jakarta	Tambora	Pekojan	0.78	35262.21	1 / S	T
Central Jakarta	New Johar	High ground	0.62	73374.54	2 / T	T

The cluster results produced by K-Means compared to the classification of population density based on the Ministry of PUPR resulted in an accuracy rate of

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

53%. The biggest cluster difference is in cluster 2 or Medium. More details can be seen in the following table.

Table 5. Matrix Comparison of Density Categories: PUPR and Laster K Results

Based on PUPR / Cluster Results	T	S	R
T	35	0	0
S	69	0	0
R	7	50	106

From the 6 examples in table 4, there is indeed an uneven population which is then divided into cluster 0, cluster 1, and cluster 2. Therefore, to form an even population, this can be a reference for the provincial government not to add residents to densely populated cluster areas. The DKI Jakarta government needs to make policies to support equal distribution of population in the DKI Jakarta area.

4. CONCLUSION

The clustering technique carried out in this study went through several stages ranging from data collection, data understanding, data processing to improve data quality and attribute selection, cluster tendency assessment, cluster formation, to cluster analysis. Clustering with the K-Means method carried out on areas in DKI Jakarta Province on population density and area data per district resulted in three regional clusters with cluster 0 (not densely populated), cluster 1 (medium population density), and cluster 2 (densely populated). These results can be used as reference material and recommendations for stakeholders in decision making in managing the distribution of population in the DKI Jakarta area as well as population management in general. The policies made are expected to support the equitable distribution of the population can be made based on the results of this analysis, besides that it is also expected to support other aspects of population management.

REFERENCES

[1] "Berapa Kepadatan Penduduk DKI Jakarta Saat Ini? - Unit Pengelola Statistik." https://statistik.jakarta.go.id/berapa-kepadatan-penduduk-dki-jakarta-saat-ini/ (accessed Oct. 17, 2021).

[2] W. Moerti, "Kepadatan Penduduk Jakarta 118 kali Lipat dari Rata-Rata Nasional," Merdeka.com, Nov. 29, 2020. https://www.merdeka.com/jakarta/kepadatan-penduduk-jakata-118-kali-lipat-dari-rata-rata-nasional.html (accessed Oct. 17, 2021).

[3] Badan Pusat Statistik Provinsi DKI Jakarta, "Pembagian Daerah Administrasi Menurut Kabupaten/Kota Administrasi di Provinsi DKI Jakarta 2019-2020," jakarta.bps.go.id, 2020. https://jakarta.bps.go.id/indicator/153/369/1/pembagian-

daerah-administrasi-menurut-kabupaten-kota-administrasi-di-provinsi-dki-jakarta.html (accessed Oct. 17, 2021).
[4] InfoJabodetabek.com, "Daftar 6 Kabupaten/Kota di Provinsi DKI Jakarta," https://jakarta.bps.go.id/. https://www.infojabodetabek.com/daftar-kabupaten-kota-di-provinsi-dki-jakarta/ (accessed Oct. 17, 2021).
[5] Republik Indonesia, "Undang-Undang No. 24 Tahun 2013 Tentang Perubahan atas Undang-Undang No. 23 Tahun 2006 Tentang Administrasi Kependudukan," 2013.
[6] R. G. Terang and R. Syafriharti, "Karakteristik Pergerakan Berdasarkan Kepadatan Penduduk Untuk Tujuan Bekerja Di Kota Bandung," Jurnal Wilayah dan Kota, vol. 7, no. 1, pp. 1–9, 2020, Accessed: Jul. 26, 2022.
[7] Menteri Pekerjaan Umum dan Perumahan Rakyat Republik Indonesia, "Peraturan Menteri Pekerjaan Umum dan Perumahan Rakyat Republik Indonesia Nomor 14/PRT/M/2018 tentang Pencegahan dan Peningkatan Kualitas Terhadap Perumahan Kumuh dan Pemukiman Kumuh," 2018. [Online]. Available: www.peraturan.go.id
[8] U. Pawitro, "Pembangunan Kota, Ekonomi Perkotaan dan Pembentukan Cluster Ekonomi Kawasan Perkotaan," in Simposium Nasional RAPI XII, 2013, pp. A1–A8.
[9] N. Pratiwi, D. B. Santosa, and K. Ashar, "Analisis Implementasi Pembangunan Berkelanjutan Di Jawa Timur," Jurnal Ilmu Ekonomi Dan Pembangunan (JIEP), vol. 18, no. 1, pp. 1–13, 2018.
[10] C. Hennig, M. Marina, M. Fionn, and R. Roberto, Handbook of Cluster Analysis. CRC Press, 2015.
[11] C. Astria, A. P. Windarto, A. Wanto, and E. Irawan, "Metode K-Means Pada Pengelompokan Wilayah Pendistribusian Listrik," in Seminar Nasional Sains & Teknologi Informasi (SENSASI) 2019, Jul. 2019, pp. 306–312.
[12] A. Singh, A. Yadav, and A. Rana, "K-means with Three different Distance Metrics," International Journal of Computer Applications, vol. 67, no. 10, pp. 13–17, Apr. 2013.
[13] M. G. Sadewo, A. P. Windarto, and A. Wanto, "Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/ Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means," KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer), vol. 2, no. 1, Oct. 2018, doi: 10.30865/KOMIK.V2I1.943.
[14] K. F. Irnanda, A. P. Windarto, I. S. Damanik, and I. Gunawan, "Penerapan K-Means pada Proporsi Individu dengan Keterampilan (Teknologi Informasi dan Komunikasi) TIK Menurut Wilayah," Seminar Nasional Sains dan Teknologi Informasi (SENSASI), vol. 2, no. 1, Aug. 2019, Accessed: Jan. 25, 2022.
[15] P. Alkhairi and A. P. Windarto, "Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," Seminar Nasional Teknologi Komputer & Sains (SAINTEKS), vol. 1, no. 1, Feb. 2019, Accessed: Jan. 25, 2022.
[16] F. Nasari, C. Jhony, and M. Sianturi, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokkan Penyebaran Diare Di Kabupaten Langkat," CogITo Smart Journal, vol. 2, no. 2, pp. 108–119, Dec. 2016, doi: 10.31154/COGITO.V2I2.19.108-119.
[17] N. Rofiqo, A. P. Windarto, and D. Hartama, "Penerapan Clustering Pada Penduduk Yang Mempunyai Keluhan Kesehatan Dengan Datamining K-Means," KOMIK

(Konferensi Nasional Teknologi Informasi dan Komputer), vol. 2, no. 1, Oct. 2018, Accessed: Jan. 25, 2022.

[18] W. Aprianti and U. Maliha, "Sistem Informasi Kepadatan Penduduk Kelurahan Atau Desa Studi Kasus Pada Kecamatan Bati-Bati Kabupaten Tanah Laut," Jurnal Sains dan Informatika, vol. 2, no. 1, pp. 21–28, 2016, Accessed: Jun. 10, 2022. [Online]. Available: https://jsi.politala.ac.id/index.php/JSI/article/view/14

[19] S. Wulandari, "Clustering Kecamatan Di Kota Bandung Berdasarkan Indikator Jumlah Penduduk Dengan Menggunakan Algoritma K-Means," in Semnas Ristek (Seminar Nasional Riset dan Inovasi Teknologi), Jan. 2020, vol. 4, no. 1. doi: 10.30998/SEMNASRISTEK.V4I1.1688.

[20] D. Gultom, I. Gunawan, I. Purnamasari, S. R. Andani, and Z. A. Siregar, "Penerapan Algoritma K-Means Dalam Pengelompokan Kepadatan Penduduk Menurut Kecamatan di Kabupaten Simalungun," Terapan Informatika Nusantara, vol. 2, no. 10, pp. 622–628, Mar. 2022, doi: 10.47065/TIN.V2I10.1375.

[21] P. Marpaung and R. F. Siahaan, "Penerapan Algoritma K-Means Clustering Untuk Pemetaan Kepadatan Penduduk Berdasarkan Jumlah Penduduk Kota Medan," Jurnal Sains Komputer dan Informatika, vol. 5, no. 1, pp. 503–521, Mar. 2021, doi: 10.30645/J-SAKTI.V5I1.343.

[22] S. Ashari, S. Khansa, C. H. M. Surudin, and I. N. Isnainiyah, "Klustering Jumlah Penduduk Kota Bandung Berdasarkan Jenis Kelamin Per Kecamatan Pada Tahun 2012 Dengan Metode K-Means," in Seminar Nasional Informatika, Sistem Informasi Dan Keamanan Siber (SEINASI-KESI), Dec. 2018, pp. 22–28.

[23] S. Sonang, A. T. Purba, and F. O. I. Pardede, "Pengelompokan Jumlah Penduduk Berdasarkan Kategori Usia Dengan Metode K-Means," Jurnal Tekinkom (Teknik Informasi dan Komputer), vol. 2, no. 2, pp. 166–172, Dec. 2019, doi: 10.37600/TEKINKOM.V2I2.115.

[24] A. S. Ahmar, D. Napitupulu, R. Rahim, R. Hidayat, Y. Sonatha, and M. Azmi, "Using K-Means Clustering to Cluster Provinces in Indonesia," Journal of Physics: Conference Series, vol. 1028, no. 1, p. 012006, Jun. 2018, doi: 10.1088/1742-6596/1028/1/012006.

[25] "Data Mining: Penerapan Algoritma K-Means Clustering dan K-Medoids... - Google Books." https://www.google.co.id/books/edition/Data_Mining_Penerapan_Algoritma_K_Means/wQnhDwAAQBAJ?hl=en&gbpv=0 (accessed Jan. 25, 2022).

[26] T. M. Kodinariya and P. R. Makwana, "Review on determining number of Cluster in K-Means Clustering," International Journal of Advance Research in Computer Science and Management Studies, vol. 1, no. 6, pp. 90–95, 2013.

[27] Mr. Dibya Jyoti Bora, A. K. Dr. Gupta, Mr. Dibya Jyoti Bora, and A. K. Dr. Gupta, "Effect of Different Distance Measures on the Performance of K-Means Algorithm: An Experimental Study in Matlab," arXiv, p. arXiv:1405.7471, May 2014, Accessed: Jan. 25, 2022.

[28] D. T. Larose and C. D. Larose, Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley & Sons, Inc, 2014.

[29] P. Makwana, T. M. Kodinariya, and P. R. Makwana, "Review on Determining of Cluster in K-means Clustering Review on determining number of Cluster in K-Means Clustering," International Journal of Advance Research in Computer Science and Management Studies, vol. 1, no. 6, 2013.