BAREKENG: Journal of Mathematics and Its Applications
June 2024 Volume 18 Issue 2 Page 1095–1104
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss2pp1095-1104

CLUSTERING ANALYSIS FOR GROUPING SUB-DISTRICTS IN BOJONEGORO DISTRICT WITH THE K-MEANS METHOD WITH A VARIETY OF APPROACHES

Denny Nurdiansyah[1*], Mochamad Nizar Palefi Ma'ady[2], Yuana Sukmawaty[3], Muchammad Chandra Cahyo Utomo[4], Tia Mutiani[5]

[1,5]Department of Statistics, Faculty of Science and Technology, Universitas Nadlatul Ulama Sunan Giri
Jln. Ahmad Yani No.10, Jambean, Sukorejo, Bojonegoro District, Surabaya, 62115, Indonesia
[2]Department of Information System, School of Industrial Engineering,
Institut Teknologi Telkom Surabaya
Jln. Ketintang No.156, Ketintang, Gayungan Sub-District, Surabaya, Jawa Timur 60231, Indonesia
[3]Department of Statistics, School of Industrial Engineering, Universitas Lambung Mangkurat
Jln. Ahmad Yani KM. 36, Loktabat Selatan, Banjar Baru Selatan, Banjar Baru, 70714, Indonesia
[4]Informatics, Department of Mathematics and Information Technology, Institut Teknologi Kalimantan
Jln. Soekarno-Hatta Km. 15, Karang Joang, Balikpapan, 76127, Indonesia

*Corresponding author's e-mail: ** **denny.nur@unugiri.ac.id**

ABSTRACT

Article History:
Received: 1st January 2024
Revised: 1st February 2024
Accepted: 24th March 2024
Published: 1st June 2024

Keywords:
Fast K-Means;
Kernel K-Means;
Population Document.

Population data is an important piece of information that is useful for regional planning and development. Insight into the state of an area is more straightforward to observe if there are grouped sub-districts. In this case, data mining techniques can identify patterns and relationships in population data. The K-Means algorithm is a clustering technique that divides data into groups or clusters based on similar characteristics. This research aims to apply the K-Means method with various approaches to clustering sub-districts in the Bojonegoro district according to population data. The research method used is a quantitative method with an exploratory study in the application of the K-Means method with a variety of approaches, namely the use of the Kernel K-Means method by utilizing the mapping function to map data to a higher dimension before the clustering process. In addition, the Fast K-Means method is used, which reduces the model training time to improve the cluster-centered recalibration problem as the amount of data increases. The data source used in this research is secondary population data in the form of birth, death, migrant, and moving variables obtained from the Satu Data Bojonegoro website developed by the Bojonegoro Regency Government. It is found that the best K-Means approach is the Kernel K-Means method with a number of clusters of 5. The performance of the cluster method is evaluated by measuring the average distance within the cluster. The data coordinate pattern in the Kernel K-means method clustering shows a smooth initial trend when the value of the number of clusters is 5 so that the clusters formed are obtained clearly. The conclusion from this study's results is that the K-Means method's best approach in grouping sub-districts in Bojonegoro district is the Kernel K-Means approach.

How to cite this article:

D. Nurdiansyah, M. N. P. Ma'ady, Y. Sukmawaty, M. C. C. Utomo and T. Mutiani, "CLUSTERING ANALYSIS FOR GROUPING SUB-DISTRICTS IN BOJONEGORO DISTRICT WITH THE K-MEANS METHOD WITH A VARIETY OF APPROACHES", *BAREKENG: J. Math. & App.*, vol. 18, iss. 2, pp. 1095-1104, June, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article • Open Access

1. INTRODUCTION

The population database process is the authority of the local government and central governments in the context of orderly population management needed by all citizens so that the data collection model established by the government is carried out to form a single citizen identification system based on systematic efforts through government issued policies that can be accessed by everyone [1]. One population data set recorded by the census and important bureau include birth, death, movement, and visitor variables from a single data site [2].

Birth is the result of the actual procreation of a woman or group of women. At the same time, death is the state of permanent loss of all signs of life, usually occurring at any time after a live birth [3]. Population arrival (migrants) is increasing the population from one place to another. In comparison, occupation (moving) decreases the population leaving or moving from one place to another [4].

The population of Bojonegoro District in 2016 was 1,306,693 people; in 2017, it increased to 1,310,079 people; in 2018, it also increased to 1,311,042 people and in 2019, it increased by 1,331,077 people; so that the population growth process of Bojonegoro District in 2020 increased by 1,344.03 8 people whose population increased due to birth, migration and economic growth [5].

The service process in Bojonegoro District is centered at the Public Service Mall on Veteran Street in Bojonegoro, where the quality of Dispendukcapil housing management services has remained the same. In Bojonegoro, the information system used is SIMDUK. This application is capable of supporting various population registration work in Bojonegoro District. Still, among the many advantages of the SIMDUK application, there are several obstacles, namely the dissemination of information about the application, which makes this application less effective. Need to achieve the maximum goal of application implementation [6]. SIMDUK is a decision that has been determined and implemented in various regions in the district/city, where this application aims to handle population problems. The data management used includes Family Cards (KK), Identity Cards (KTP), Birth Certificates (AK), Population Census, and Population Demographics [6].

However, this information system has no longer been used after COVID-19. Now, the village head relies only on data from the population status office, which is taken once a year as an Excel file. The data is entered once a year, and public awareness of population registration still needs to be higher. As a result, the population data each year must be more accurate, and this problem repeats itself yearly. Village officials must start checking the data manually every year [7].

Data mining is known as knowledge discovery in databases (KDD). It is an activity that involves collecting and using historical data to discover relational rules and patterns in databases of large data sets and is one of the statistical methods for analyzing data [8]. A technique to put data into groups or clusters with a particular affinity for each object is clustering. Generally, this clustering concept classifies a set of objects into several groups without deep knowledge about the groups [9]. The primary purpose of clustering is to classify a data set into clusters with similar characteristics and between clusters with different characteristics. One of the clustering techniques is the K-Means technique [10].

K-Means is a clustering method that uses the concept of partitioning, which, later in the process, the algorithm divides the data into many different clusters or groups. By partitioning iteratively, K-Means can reduce the average distance or distance of data per cluster. The essence of K-Means is a method of applying clustering without any direction (unsupervised) [11]. In the comparative study of K-Means, K-Medoids, and Random Clustering, the K-Means Method is the best method in clustering elementary schools in Bojonegoro District of 723 schools based on capacity and facilities [12]. In the clustering study with modification of the Elbow Method, the best method in clustering sub-districts in Bojonegoro District of 28 sub-districts based on several types of assistance aid is the K-Means Method [13]. The K-Means algorithm has several limitations, namely random initialization of the centroid, which can lead to unexpected convergence, requires determining the number of clusters beforehand, the shape of the clusters can be different due to the effect of outliers, and its inability to handle various types of data [14].

One of the popular clustering algorithms is the K-Means method. Besides its simplicity, it is also able to provide efficient results. The disadvantage of this algorithm is that the classifier formed must be linearly separable, so K-Means only works well when experiencing data problems with nonlinear decomposition conditions. Hence, it needs to be expanded in size mapped by a kernel. The kernel is used as a variable whose class is searched by K-Means or called Kernel K-Means (KKC) so that K-Means works by determining the position of the center randomly in advance and based on previous searches; the randomization affects the cluster results themselves, where the results are unstable. In some cases, it can also lead to suboptimal results [15]. A representative kernel needs to be generated from the Kernel K-Means method by optimizing the combination of weights given in the clustering process [16].

The Fast K-Means Method is an algorithm that shortens the image database's cluster center formation time. It solves the cluster center recycling problem as many images are continuously added to the color image database for performance comparison and analysis [17]. The problem is shortening the time spent on training cluster centers and addressing the issue of retraining cluster centers, as many points are continuously added to the dataset [18]. The standard K-Means algorithm is very slow when clustering millions of data into thousands of groups or tens of thousands of clusters, so this Fast K-Means is an alternative solution [19]. The speed problem in question is how quickly to get neighbor information by reducing the scale of distance calculations and neighbor update strategies so that neighbor selection is more accurate for each cluster at each iteration until a convergent condition is reached [20].

Based on previous research, this study will compare variations of K-Means methods: Standard K-Means, Kernel K-Means, and Fast K-Means. This comparative study of K-Means variants is new in this research, and population data, including birth, death, migrant, and moving, is used. This method is applied using Rapidminer software.

The purpose is to apply the K-Means method with various approaches to group the sub-districts in the Bojonegoro district based on population data. Performed Comparison of several K-Means methods, namely Standard K-Means, Kernel K-Means, and Fast K-Means on population data, so that the results of clustering sub-districts in Bojonegoro Regency based on population data were found with the best K-Means method. In addition, this study also displays the results of descriptive statistics of the clusters formed. This research is helpful for the Bojonegoro Population and Civil Registry Office as for the benefits obtained from this research, the Bojonegoro District Population and Civil Registry Office can input material in the form of grouping knowledge so that it becomes a consideration in taking the next step in making policies related to public services, especially in the administration of population documents.

Statistical analysis, especially the K-Means method, is suitable for analyzing sub-district clustering data because it can identify similar demographic patterns based on the birth, death, migrant, and moving variables. The use of this method allows researchers to efficiently understand the complex interactions between these variables in the formation of sub-district groups. The results can provide important insights for developing regional policy decision-making and population management. Grouping sub-districts in Bojonegoro District based on population data using clustering can provide input in the form of sub-district groupings that help data collection to determine population control goals and poverty rates. Therefore, it is necessary to propose this study.

2. RESEARCH METHODS

2.1 Data Source

The data source used in this research is secondary population data form of birth, death, migrant, and moving variables obtained from the Satu Data Bojonegoro website in 2022. The Bojonegoro District Government developed this website.

2.2 Research Variables

In this study, the research variables will be presented for response and predictor variables in Table 1 as follows:

Table 1. Definition of Research Variables for Response and Predictor Variables

Attributes	Variable Name	Definition	Measurement Scale
Label	**Sub-district**	The sub-districts in Bojonegoro District are Balen, Baureno, Bojonegoro, Bubulan, Dander, Gayam, Gondang, Kalitidu, Kanor, Kapas, Kasiman, Kedewan, Kedungadem, Kepohbaru, Malo, Margomulyo, Ngambon, Ngasem, Ngraho,	Nominal

Attributes	Variable Name	Definition	Measurement Scale
		Padangan, Purwosari, sekar, Sugihwaras, Sukosewu, Sumberrejo, Tambakrejo, Temayang, and Trucuk.	
Common	**Births**	Total birth variable in each sub-district.	Ratio
	Death	Total death variable in each sub-district.	Ratio
	Migrant	Total of newcomers variable in each sub-district.	Ratio
	Moving	Total people who moved out variable in each sub-district.	Ratio

2.3 Data Analysis

The data analysis in this study is a grouping with the K-means method with variations of approaches: Standard K-means, Kernel K-Means, and Fast K-Means, using the Rapidminer Studio that offers a Free Plan with limited features. Rapidminer Studio also offers an Enterprise Plan for the product. The following are the research procedures to be carried out [14].

1. Prepare population administration data containing case labels and other variables used as references in the clustering analysis.
2. Determine the characteristics of population administration data using descriptive statistical analysis.
3. Evaluate the Standard K-means, Kernel K-Means, and Fast K-Means methods to obtain a Performance Vector containing the average distance and each cluster's average distance.
4. Displaying clusters graphically for Standard K-means, Kernel K-Means, and Fast K-Means methods.
5. Comparing the Performance Vector results, namely the analysis results' value per class from $k = 2$ to $k = 10$, to determine the best method based on the minor average within cluster distance.
6. Performing regional grouping of each sub-district based on the results of the best clustering method to determine the level of birth, death, migrant, and moving in Bojonegoro District.
7. Display descriptive statistics for each cluster.

The Standard K-Means data processing steps taken from the journal [15]:

1. Determine the value of k as the cluster formed.
2. Initialization of k cluster centers can be done in various ways. Still, the most common is randomly drawing from existing data.
3. Calculate the distance to each centroid for all input data using the Euclidean distance formula until the closest distance between each data and centroid is found.
4. Grouping each data based on its proximity to each centroid.
5. Update the centroid value iteratively; the new centroid value is obtained from the average of the clusters formed.
6. Repeating steps 2 to 5 until there are no different cluster members.

The Kernel K-Means data is done with processing steps taken from the journal [16] which are as follows.

1. Entering data for descriptive statistics processing.
2. Standardizing the data.
3. Performing KMO test.
4. Performing multicollinearity check.
5. Input the number of clusters, or $K = 2,3,4,5$, with cluster initials C_1, C_2, C_3, and C_4.
6. Input Kernel matrix K of size $N \times N$.
7. Calculating the initial center (m_k)
8. Calculates the new distance (δ_{kn}) for all values of n at all cluster centers (m_k).
9. Determines the value of $C^*(X_n)$ to decide the nth data at all cluster centers ($C(k)$) with the closest distance.
10. Update $C(k)=\{\{X(n)\}|C^*(X(n)) = k\}$ until all values of $C(k)$ converge.
11. Calculating the validity index.
12. Determine the best number of clusters from the cluster validity index values.
13. Interpret the formed cluster profile.

The Fast K-Means is done with data processing steps given in the following procedure [17]:

1. Partition the data points into k clusters, Si (i = 1,2...., k). Cluster Si will be associated with a representative cluster center (Ci).
2. Define the set of data points as S = {X}
3. Suppose d{X, Y} is the distortion between two vectors X and Y, where d{X, Y} is the Euclidean distance between X and Y.
4. Suppose the centers of the first cluster in the current and previous partitions are used with the notations Ci and Ci'.
5. Define the displacement between Ci and Ci' as Di, if Di = ||Ci - Ci'||. If Di = 0, then the vector Ci is defined as the static and active cluster center where the cluster is associated with the active cluster center.

A flowchart is a representation of the algorithm in a decision-making process with the selection of the best method in the Clustering Study or grouping with the K-Means Variance method, which includes Standard K-means, Kernel K-Means, and Fast K-Means. In this study, the resulting data analysis stages follow the appropriate flow chart in **Figure 1** below:

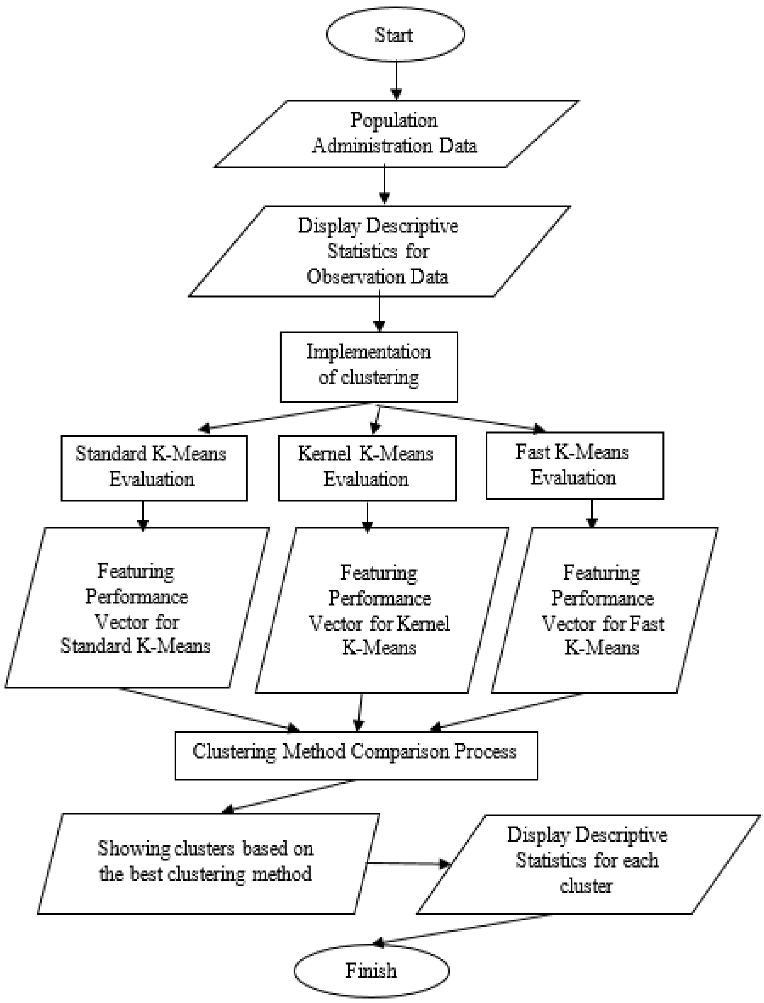

Figure 1. Flowchart of Data Analysis Steps.

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistics

In this study, the amount of data used is 28 data. All of the sub-districts in Bojonegoro District regarding population data in the form of birth, death, migrant, and moving variables in 2022. The data was collected on February 27, 2023, based on data obtained on the Bojonegoro One data website as research data for research materials, which were then converted into an Excel file. Population data was taken in the vulnerable months of January - December 2022. In this study, descriptive statistics were obtained with the help of SPSS version 24 software, which is presented in **Table 2** below:

Table 2. Descriptive Statistics of Population Data.

Variable	Number of Data	Minimum	Maximum	Sum	Mean	Standard Deviation
Birth	28	127	1,133	15,418	550.64	279.266
Death	28	69	1,051	13,979	499.25	278.476
Migrant	28	121	1,585	14,565	520.18	314.362
Moving	28	115	1,340	13,507	482.39	282.327

Table 2 summarizes descriptive statistics for grouping population data based on birth, death, migrant, and moving variables in 28 sub-districts in Bojonegoro district. The lowest birth value is 127 in the Ngambon sub-district, while the highest is in the Dander sub-district. The number of births in Bojonegoro district is 15,418, with a mean value of 550.64 and a standard deviation of 279.266. Furthermore, the second variable is death, which has the lowest value of 69 in the Gondang sub-district, and the highest death value is in the Bojonegoro sub-district. For the total number of deaths, as many as 13,979 people, with an average of 499.25 and a standard deviation of 278.476.

The third variable is migrant, with the lowest value of 121 in the Ngambon sub-district and the highest moving value of 1,585 in the Bojonegoro sub-district. The total number of moves amounted to 14,565 residents, with an average of 520.18 and a standard deviation of 314.362. The last population data variable is moving, with the lowest number of moving at 115 in the Ngambon sub-district and the highest number of moving at 1,340 in the Bojonegoro sub-district, for the total number of moving data as many as 13,507 with a mean value of 482.39 and a standard deviation of 282.327.

3.2 K-Means Method with a Variety of Approaches

In grouping sub-districts in Bojonegoro District based on population data, we can know the results of data processing from Rapidminer Studio software, which contains the third variance of K-Means as follows:

Table 3. Results of Average Within Distance from Implementing The K-Means Method of The Three Approaches.

Group	Average Within Distance		
	Standard K-Means Method	**Kernel K-Means Method**	**Fast K-Means Method**
$k = 2$	-18.803	-18.053	-18.803
$k = 3$	-12.769	-9.867	-12.769
$k = 4$	-5.729	-9.867	-7.729
$k = 5$	-4.549	-3.939	-4.549
$k = 6$	-2.807	-3.939	-4.184
$k = 7$	-2.405	-2.522	-2.405
$k = 8$	-2.084	-2.950	-1.995
$k = 9$	-1.705	-2.593	-1.393

Based **Table 3** shows the goodness of fit of the Average Within Distance method. In the Elbow method, the most optimal group k value is obtained from observing the first sloping line diagram at a specific k value. This criterion follows the approach of the Elbow method **[13]**. The following visualizations of the line diagrams for the three approaches of the K-Means method are given in **Figure 2** below.

Figure 2. Line Chart for Average Within Distance for the K-Means Method of the Three Approaches.

Figure 2 shows that the first ramping point of the three approaches for the K-Means method falls at k of 5, which is marked by the black line. The numerical value of the Average Within Distance of the three methods can be observed in Table 3. The table shows that the best method falls on the Kernel K-Mean method with the Average Within Distance value closest to 0, which is obtained at -3.939. The closest distance value must have the smallest distance value.

3.3 District Clustering with the Best K-Means Method.

With the Kernel K-Means method as the best method, the sub-districts in Bojonegoro district will be determined as 5 groups, and their members are shown in Table 4.

Table 4. Grouping of Sub-Districts Based on Clusters.

Cluster	Sub-Districts
1	Ngraho, Sugihwaras, Sukosewu, Temayang, Trucuk.
2	Bubulan, Dander, Kedewan, Ngambon.
3	Balen, Baureno, Bojonegoro, Kanor, Kapas, Kedungadem, Kepohbaru, Ngasem, Sumberrejo.
4	Kalitidu, Padangan, Tambakrejo.
5	Gayam, Gondang, Kasiman, Malo, Margomulyo, Purwosari, Sekar.

The formed sub-district clusters cannot yet be characterized, so the usefulness of this clustering method still needs to be seen. Thus, descriptive statistics per cluster will also be provided.

3.4 Descriptive Statistics for Each Cluster.

When clusters are formed, descriptive statistics are needed so that the groups of sub-districts formed can be observed for their characteristics. Descriptive statistics for each cluster are given below:

Table 5. Descriptive Statistics for Each Cluster.

Cluster	Variable	Number Of Object	Minimum	Maximum	Sum	Mean	Standard Deviation	Ranking
1	**Birth**	5	425	586	2492	498.40	61.647	3
	Death	5	430	517	2355	471.00	34.950	3
	Migrant	5	332	520	2111	422.20	67.736	3
	Moving	5	366	450	2060	412.00	33.226	3

Cluster	Variable	Number Of Object	Minimum	Maximum	Sum	Mean	Standard Deviation	Ranking
2	**Birth**	4	127	1133	1586	396.50	491.446	4
	Death	4	121	1052	1453	363.25	459.396	4
	Migrant	4	109	1032	1498	374.50	441.123	4
	Moving	4	115	1012	1424	356.00	437.634	4
3	**Birth**	9	671	1008	7418	824.22	116.128	1
	Death	9	577	1585	7245	805.00	306.787	1
	Migrant	9	575	1051	7018	779.78	137.640	1
	Moving	9	546	1340	6706	745.11	238.361	1
4	**Birth**	3	498	658	1801	600.33	88.861	2
	Death	3	504	570	1590	530.00	35.157	2
	Migrant	3	522	632	1690	563.33	59.878	2
	Moving	3	490	564	1587	529.00	37.162	2
5	**Birth**	7	223	410	2121	303.00	62.303	5
	Death	7	220	352	1922	274.57	43.855	5
	Migrant	7	69	307	1662	237.43	82.259	5
	Moving	7	184	316	1730	247.14	44.861	5

Table 4 and **Table 5** show that there are 5 sub-districts in Cluster 1, namely Ngraho, Sugihwaras, Sukosewu, Temayang, and Trucuk. Cluster 1 is a group of sub-districts that have Birth, Death, Migrant, and Moving values in the third rank in Bojonegoro district, with each sub-district's average value of 498.40, 471.00, 422.20, and 412.00. In cluster 2, there are four sub-districts: Bubulan, Dander, Kedewan, and Ngambon. Cluster 2 is a group of sub-districts that have the second lowest Birth, Death, Migrant, and Moving values in the Bojonegoro district, with each sub-district's average value of 396.50, 363.25, 374.50, and 356.00. For cluster 3, there are nine sub-districts: Balen, Baureno, Bojonegoro, Kanor, Kapas, Kedungadem, Kepohbaru, Ngasem, and Sumberrejo. Cluster 3 is a group of sub-districts that have the highest Birth, Death, Migrant, and Moving values in Bojonegoro district, with each sub-district's average value of 824.22, 805.00, 779.78, and 745.11. In cluster 4, there are three sub-districts: Kalitidu, Padangan, and Tambakrejo. Cluster 4 is a group of sub-districts that have the second highest Birth, Death, Migrant, and Moving values in the Bojonegoro district, with each sub-district's average value of 600.33, 530.00, 563.33, and 529.00. Meanwhile, cluster 5 consists of seven sub-districts, including Gondang, Gayam, Kasiman, Malo, Margomulyo, Purwosari, and Sekar. Cluster 5 is a group of sub-districts that have the lowest Birth, Death, Migrant, and Moving values in the Bojonegoro district, with each sub-district average value of 303.00, 274.57, 237.43, and 247.14.

Based on the data analysis from **Table 4** and **Table 5**, the recommendation for Bojonegoro district is the subsequent research that should conducted to understand the factors behind the ranking of sub-districts in certain clusters. It is necessary to evaluate strategies or policies that have been successful in low-scoring clusters to guide other areas. In addition, there is a need to focus on addressing health or social issues in high-scoring clusters, identifying the reasons behind high scores in certain clusters, and finding solutions to optimize the condition of these areas.

4. CONCLUSIONS

From the results of the comparison of catering methods on the application of population data, it can be seen that each approach has its ramps based on the value of the group (cluster) and the results of the calculation of the average within-distance value of each K-Means variation with k of 5. The kernel K-Means method is the best approach. Cluster 3 is the sub-district group that ranks the highest in population data, while Cluster 5 is the lowest. Suggestions for further research related to research data are expected to be grouped at the village level. Developing the K-Mean Kernel approach needs to be tried on other types of Kernel.

ACKNOWLEDGMENT

We thank Universitas Nahdlatul Ulama Sunan Giri (UNUGIRI) for granting us access to the computer lab necessary to conduct this study. We would also like to thank the Bojonegoro District Government for providing the Satu Data website, which has helped provide population data information so researchers can access it directly.

REFERENCES

[1] U. U. Muhimah, "Peran Pemerintah dalam Bidang Administrasi Kependudukan dalam Kerangka Perlindungan Hukum Warga Negara Ditinjau dari Undang-Undang No. 23 Tahun 2006 tentang Administrasi Kependudukan," *Sultan Jurisprud. J. Ris. Ilmu Huk.*, vol. 2, no. 1, pp. 53–63, 2022, doi: 10.51825/sjp.v2i1.15879.

[2] N. A. Wardrop *et al.*, "Spatially disaggregated population estimates in the absence of national population and housing census data," *Proc. Natl. Acad. Sci.*, vol. 115, no. 14, pp. 3529–3537, Apr. 2018, doi: 10.1073/pnas.1715305115.

[3] K. D. Negeri, *Peraturan Menteri Dalam Negeri Republik Indonesia Nomor 40 Tahun 2012 Tentang Pedoman Penyusunan Proyeksi Penduduk Di Daerah*. 2012, pp. 1–9.

[4] O. Y. Fujiyati and Sukadi, "Sistem Informasi Pengolahan Data Kependudukan Desa Purwoasri," *J. Speed – Sentra Penelit. Eng. dan Edukasi*, vol. 7, no. 1, pp. 1–15, 2015.

[5] W. P. Bojonegoro, "Demografi Kabupaten Bojonegoro," *Dinas Kependud. dan Pencatatan Sipil*, 2019.

[6] N. A. Sholikhah, "Studi Perbandingan Clustering Kecamatan di Kabupaten Bojonegoro Berdasarkan Keaktifan Penduduk Dalam Kepemilikan Dokumen Kependudukan," *J. Stat. dan Komputasi*, vol. 1, no. 1, pp. 42–53, Jun. 2022, doi: 10.32665/statkom.v1i1.443.

[7] I. G. I. Sudipa and E. A. P. Lestari, "Rancang Bangun Sistem Informasi Penduduk Dusun (Studi Kasus : Dusun Tegal Kori Kaja Ubung)," *J. Teknol. Inf. dan Komput.*, vol. 5, no. 2, 2019, doi: 10.36002/jutik.v5i2.782.

[8] H. A. Madni, Z. Anwar, and M. A. Shah, "Data mining techniques and applications — A decade review," in *2017 23rd International Conference on Automation and Computing (ICAC)*, Sep. 2017, pp. 1–7. doi: 10.23919/IConAC.2017.8082090.

[9] G. Ahalya and H. M. Pandey, "Data clustering approaches survey and analysis," in *2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)*, Feb. 2015, pp. 532–537. doi: 10.1109/ABLAZE.2015.7154919.

[10] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J*, vol. 2, no. 2, pp. 226–235, Jun. 2019, doi: 10.3390/j2020016.

[11] K. F. Irnanda, D. Hartama, and A. P. Windarto, "Analisa Klasifikasi C4.5 Terhadap Faktor Penyebab Menurunnya Prestasi Belajar Mahasiswa Pada Masa Pandemi," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 327, 2021, doi: 10.30865/mib.v5i1.2763.

[12] D. Nurdiansyah, S. Saidah, and N. Cahyani, "DATA MINING STUDY FOR GROUPING ELEMENTARY SCHOOLS IN BOJONEGORO DISTRICT BASED ON CAPACITY AND EDUCATIONAL FACILITIES," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 2, pp. 1081–1092, Jun. 2023, doi: 10.30598/barekengvol17iss2pp1081-1092.

[13] H. Fitriyah, E. M. Safitri, N. Muna, M. Khasanah, D. A. Aprilia, and D. Nurdiansyah, "IMPLEMENTASI ALGORITMA CLUSTERING DENGAN MODIFIKASI METODE ELBOW UNTUK MENDUKUNG STRATEGI PEMERATAAN BANTUAN SOSIAL DI KABUPATEN BOJONEGORO," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. Dan Stat.*, vol. 4, no. 3, pp. 1598–1607, 2023, doi: https://doi.org/10.46306/lb.v4i3.453.

[14] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation," *Electronics*, vol. 9, no. 8, p. 1295, Aug. 2020, doi: 10.3390/electronics9081295.

[15] K. Aprianto, "Optimasi Kernel K-Means dalam Pengelompokan Kabupaten/Kota Berdasarkan Indeks Pembangunan Manusia di Indonesia," *Limits J. Math. Its Appl.*, vol. 15, no. 1, p. 1, 2018, doi: 10.12962/limits.v15i1.3408.

[16] Y. Yao, Y. Li, B. Jiang, and H. Chen, "Multiple Kernel k -Means Clustering by Selecting Representative Kernels," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 32, no. 11, pp. 4983–4996, Nov. 2021, doi: 10.1109/TNNLS.2020.3026532.

[17] A. N. Azizah, T. Widiharih, and A. R. Hakim, "Kernel K-Means Clustering untuk Pengelompokan Sungai di Kota Semarang Berdasarkan Faktor Pencemaran Air," *J. Gaussian*, vol. 11, no. 2, pp. 228–236, 2022, doi: 10.14710/j.gauss.v11i2.35470.

[18] C.-H. Lin, C.-C. Chen, H.-L. Lee, and J.-R. Liao, "Fast K-means algorithm based on a level histogram for image retrieval," *Expert Syst. Appl.*, vol. 41, no. 7, pp. 3276–3283, Jun. 2014, doi: 10.1016/j.eswa.2013.11.017.

[19] Q. Hu, J. Wu, L. Bai, Y. Zhang, and J. Cheng, "Fast K-means for Large Scale Clustering," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Nov. 2017, pp. 2099–2102. doi: 10.1145/3132847.3133091.

[20] D. Peng, Z. Chen, J. Fu, S. Xia, and Q. Wen, "Fast k-means Clustering Based on the Neighbor Information," in *2021*

International Symposium on Electrical, Electronics and Information Engineering, Feb. 2021, pp. 551–555. doi: 10.1145/3459104.3459194.