Jurnal Pendidikan Matematika (Kudus)
P-ISSN 2615-3939 | E-ISSN 2723-1186
https://journal.iainkudus.ac.id/index.php/jmtk
DOI: http://dx.doi.org/10.21043/jpmk.v5i2.16552
Volume 5, Number 2, December 2022, pp. 185-198

Comparison of K-Means, Fuzzy C-Means, Fuzzy Gustafson Kessel, and DBSCAN for Village Grouping in Surabaya Based on Poverty Indicators

Sri Hidayati[1*], Aviolla Terza Darmaliana[2], Raulia Riski[1]
[1]Department of Information System, Institut Teknologi Telkom Surabaya, Surabaya, Indonesia
[2]Departement of Data Science, Universitas Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia
** Correspondence: srihidayati@ittelkom-sby.ac.id*

Abstract

The population growth rate in various countries in the world is increasing, including Indonesia. The population explosion as a result of rapid population growth has a negative impact on the socio-economic life of the community, such as increasing unemployment rates, food shortages, and high poverty rates. Therefore, local governments in each country try to overcome the poverty problem using various policies, including in Surabaya, East Java, Indonesia. This study aims to classify villages in Surabaya using non-hierarchical clustering, such as K-Means, Fuzzy C-Means, Fuzzy Gustafson Kessel, and DBSCAN (Density-Based Spatial Clustering of Applications with Noise), based on poverty indicators. Before analysis, the villages in Surabaya, East Java, Indonesia were classified using non-hierarchical clustering, and the results of cluster analysis were compared from various methods using the value of within clusters sum of squares and average silhouette width. Comparison between village grouping methods results in K-Means being the best method for village grouping in Surabaya, East Java, Indonesia based on the values of the within clusters sum of squares. While based on the average silhouette width value, the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) method is found to be the best method because its value was close to 1 compared to the other methods. Thus, it can be implicated that K-Means and DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is the best method for village grouping in Surabaya, East Java, Indonesia in relation to poverty problems.

Keywords: DBSCAN; Fuzzy C-Means; Fuzzy Gustafson Kessel; K-Means; Non-Hierarchical Clustering

Abstrak

Laju pertumbuhan penduduk di berbagai negara di dunia semakin meningkat, termasuk Indonesia. Ledakan penduduk akibat pertumbuhan penduduk yang pesat berdampak negatif terhadap kehidupan sosial ekonomi masyarakat, seperti meningkatnya angka pengangguran, kekurangan pangan, dan tingginya angka kemiskinan. Oleh karena itu, pemerintah daerah di setiap negara berusaha mengatasi masalah kemiskinan dengan berbagai kebijakan, termasuk di Surabaya, Jawa Timur, Indonesia. Penelitian ini bertujuan untuk mengklasifikasikan desa-desa di Surabaya, Jawa Timur, Indonesia menggunakan *non-hierarchial clusterings,* seperti *K-Means, Fuzzy C-Means, Fuzzy Gustafson Kessel, dan DBSCAN (Density-Based Spatial Clustering of Applications with Noise)*, berdasarkan indikator kemiskinan. Sebelum dilakukan analisis, desa-desa di Surabaya, Jawa Timur, Indonesia diklasifikasikan menggunakan *non-hierarchical clustering*, dan hasil analisis *cluster* dibandingkan dari berbagai metode dengan menggunakan nilai cluster sum of squares dan rata-rata lebar siluet. Perbandingan antar metode pengelompokan desa menghasilkan *K-Means* menjadi metode terbaik untuk pengelompokan desa di Surabaya berdasarkan nilai *cluster sum of squares*. Sedangkan berdasarkan nilai rata-rata lebar siluet, metode *DBSCAN (Density-Based Spatial Clustering of Applications with Noise)* merupakan metode yang paling baik karena nilainya mendekati 1 dibandingkan dengan metode lainnya. Dengan demikian, dapat diimplikasikan bahwa *K-Means* dan *DBSCAN (Density-Based Spatial Clustering of Applications with Noise)* merupakan metode terbaik untuk pengelompokan desa di Surabaya, Jawa Timur, Indonesia dalam kaitannya dengan masalah kemiskinan.

Kata Kunci: DBSCAN; Fuzzy C-Means; Fuzzy Gustafson Kessel; K-Means; Pengelompokan Non-Hierarki

Introduction

The population growth rate in various countries in the world is increasing, including Indonesia, which was 1.49 percent from 2000 to 2010 (BPS, 2021). This shows that the population in Indonesia increased by around 3.5 million annually. The population explosion as a result of rapid population growth has a negative impact on the socio-economic life of the community, such as increasing unemployment, food shortages, and high poverty rates (Soegimo & Ruswanto, 2009).

Poverty is the inability to meet minimum living standards. BPS bases the relative minimum standard size on the number of rupiah spent per capita/month to meet the minimum food and non-food products needs. Minimum food requirements use a benchmark of 2,100 calories/day, while non-food needs include housing, clothing, as well as various goods and services. The government continues to strive to reduce poverty every year (BPS Provinsi Jawa Timur, 2020). Local governments in each country try to overcome the problem of poverty using various policies, including in Surabaya, East Java, Indonesia. Several programs for poverty reduction have also been carried out such as rice for poor families, BPJS health, and cash advance assistance. But these efforts have not shown significant results.

Surabaya consists of 31 sub-districts and 154 villages. Surabaya, the second largest city in Indonesia, is currently aggressively pursuing poverty alleviation. The Central Bureau of Statistics noted that the poverty rate of Surabaya City in 2017 reached 5.39 percent of the total poor population at a 155 thousand people (BPS Kota Surabaya, 2018). Target inaccuracy in poverty alleviation caused the poverty alleviation goal to be far from expectations. Therefore, various government policies related to poverty will certainly be more effectively implemented if the poverty indicators can be more precisely defined.

In terms of the poverty level, the villages in Surabaya are grouped based on the highest and the average poverty levels, then the areas with low poverty levels are expected to be facilitated by the government policies as a target to overcome poverty problems there. Non-hierarchical cluster analysis is a method that places objects into clusters at once so that certain clusters are formed. Non-hierarchical methods used in this study are K-Means, Fuzzy C-Means, Fuzzy Gustafson Kessel, and DBSCAN. These methods grouped observation objects that are close to each other based on distance measurements and the minimum number of values (Gueorguieva, Valova, & Georgiev, 2017). The non-hierarchical clustering method was also examined by several researchers, among others (Pal & Bezdek, 1995; Tilson, Excell, & Green, 1988; Li, Ng, Cheung, & Huang, 2008; Oyelade, Oladipupo, & Obagbuwa, 2010;Park, Park, Yoo, Choi, & Han, 2020; Serrao et al., 2018; Treiger, Bondarenko, van Malderen, & van Grieken, 1995; Askari, 2021). This study aims to classify villages in Surabaya using non-hierarchical clustering, namely K-Means, Fuzzy C-Means, Fuzzy Gustafson Kessel, and DBSCAN to help the government have more accurate target on applying policies to overcome poverty problems in Surabaya. Poverty indicators used, namely the percentage of households with ownership status of land based on other people's property (X_1), the percentage of households with drinking water sources from meter tap (X_2), the percentage of households with 450 watts of electricity (X_3), the percentage of households with 3 kg cooking fuel gas (X_4), the percentage of households with the widest floor type other than ceramics, marble, rugs (X_5), and the percentage of households which are Raskin program participants (X_6).

Method

Clustering Method

Clustering is a process of grouping data objects into disjointed clusters so that the data in the same cluster are similar, while at the same time they differ from the data in other clusters. In other words, a cluster is a collection of data object that are similar to one another within the same cluster and dissimilar to the objects in other clusters (Härdle & Simar, 2007). The cluster analysis procedure was used to identify groups of cases that are relatively similar based on the characteristics that have been selected using an algorithm to manage large numbers of cases. The algorithm used required the number of clusters to be created. The classification method can be used were either one of the two, namely to update cluster groups iteratively or to only classify.

To analyze the cluster, the processes need to be carried out were as follows. Stage 1, similarities between objects were measured. In accordance with the principle of cluster analysis which groups objects with similarities, the first process was to measure how far were the similarities between these objects. Stage 2, in the cluster there were hierarchical and non-hierarchical methods. The advantage of using the hierarchical method in cluster analysis were to speed up the processing and to save time, because the data inputted will form a hierarchy or form separate levels thus makes it easier to be interpreted. The disadvantages of this method are there were often errors in outlier data, there might be size differences of the distance used, and there might be irrelevant variables. On the other hand, the non-hierarchical method has the advantage of being able to analyze larger sizes samples more efficiently. In addition, there were also only a few weaknesses in the outlier data, the distance used, and the irrelevant or incorrect variables. Its weakness is the point of randomness which is worse than the hierarchical method (Härdle & Simar, 2007).

Unlike the hierarchical method, non-hierarchical method starts with the first desired number of clusters (two clusters, three clusters or the other). After the number of clusters is known, the cluster process is done without following the hierarchy process. This method is commonly referred to K-Means cluster. Another methods are Fuzzy C-means and Fuzzy Gustafson Kessel (Raval & Jani, 2016).

K-Means Method

There have been many clustering techniques proposed but K-Means is one of the oldest and most popular clustering techniques. In this method, the number of cluster (k) was predefined prior to analysis, then the selection of the initial centroids would be made randomly, and then followed by an iterative process of assigning each data point to its nearest centroid. This process will keep repeating until the convergence criteria is met (Agarwal, 2014).

According to the basic K-Means clustering algorithm, clusters were fully dependent on the selection of the initial clusters' centroids. K data elements were selected as initial centers, then the distances of all data elements were calculated by the Euclidean distance formula. Data elements with less distance to centroids were then moved to the appropriate cluster. The process was continued until no more changes occur in the clusters [k-1]. This partitioning clustering is the most popular and fundamental technique (Agusta, 2007).

Several alternative applications of K-Means with several compilations of related calculation theories have been proposed, this includes the elections. Distance space to calculate the distance between a data and a centroid and data allocation method returns to each cluster (Agusta, 2007).

Fuzzy C-Means Method

Fuzzy K-Means method (or more commonly used as Fuzzy C-Means) reallocated data into each cluster by using Fuzzy theory. This method used membership function variable, u_{ik}, which refers to how likely a data can become a member in a cluster (Jalali, 2016). In Fuzzy K-Means, a variable, m, was also introduced as the weighting exponent of the membership function. This variable could change the magnitude of the membership function effect, u_{ik}, in the clustering process using the Fuzzy K-Means method. Membership function for a data to a particular cluster was calculated using the following formula.

$$u_{ik} = \sum_{j=1}^{c} \left(\frac{D(x_k, v_i)}{D(x_k, v_j)} \right)^{\frac{2}{m-1}}$$

Annotation:

$u_{ik}: membership\ function\ of\ the\ k^{th}\ data\ to\ the\ i^{th}\ cluster$

$v_i\ :\ the\ centroid\ value\ of\ the\ i^{th}\ cluster$

$m\ :\ weighting\ exponent$

Membership function had an area of $0 \leq u_{ik} \leq 1$. Item data that has a higher likelihood level to a group will have a membership function value to the group that is closed to 1 and the other group approaches 0 (Li, Cheung & Huang, 2008).

Fuzzy Gustafson Kessel

Gustafon Kessel, proposed a modification to the distance component in the D_{ik}^2 objective function which is minimized in the FCM, namely the mahalanobis distance formula which was for hyper ellipsoidal forms and to consider the distribution of data by entering data covariance (Härdle & Simar, 2007). Matrix A was proposed to be substituted with a matrix called fuzzy covariance matrix. This caused the grouping of FGK to be able to better adjust the geometric shape of the membership function that was right for a set. These were the algorithm of FGK method.

1. Input the data that will be clustered;
2. Determine how many groups to be formed $(1 < c < N)$, weighting exponent $(m > 1)$, maximum iteration (maslter), smallest error $(\varepsilon > 0)$, initial objective function=0, and initial iteration $(t = 1)$;
3. Form the initial partition matrix U;
4. $U = \begin{bmatrix} u_{11}(x_1) & u_{12}(x_2) & \dots & u_{1I}(x_I) \\ u_{21}(x_1) & u_{22}(x_2) & \dots & u_{2I}(x_I) \\ \vdots & \vdots & \ddots & \vdots \\ u_{c1}(x_1) & u_{c2}(x_2) & \cdots & u_{cI}(x_I) \end{bmatrix}$
5. Calculate the group center to k with the following formula;
6. $v_i = \frac{\sum_{k=1}^{N} u_{ik}^m x_k}{\sum_{k=1}^{N} u_{ik}^m}, i = 1,2,\dots,c$
7. Calculate the group covariance matrix;
8. $D_{ik}^2 = \|x_k - v_i\|^2 = (x_k - v_i)^T A (x_k - v_i)$
9. Where A = positive definite matrix;
10. Calculate the objective function in equation (3) in t iteration;
11. Calculate the value of the new membership function $U^{(t+1)}$
12. $u_{ik} = \left[\sum_{j=1}^{c} \left(\frac{D(x_k, v_i)}{D(x_k, v_j)} \right)^{\frac{2}{m-1}} \right]^{-1}$
13. Comparing membership values in matrix U, where $\|U^{t+1} - U^t\| < \varepsilon$ or $(t > MaksIter)$ is converging. If $\|U^{t+1} - U^t\| \geq \varepsilon$ then goes back to step 4 (Härdle & Simar, 2007).

DBSCAN

DBSCAN or Density-Based Spatial Clustering and Application with Noise can be used to identify clusters of various form of datasets that contain noise and outliers. Some advantages of DBSCAN are it does not specify the number of clusters (forms not just circles), and this method can identify outliers. Two important parameters are needed in DBSCAN, namely epsilon and minimum points. The epsilon parameter defines the radius around point x and the minimum point parameter is the number of observations within the epsilon radius (Härdle & Simar, 2007).

Selection of The Best Method

To select the best grouping method, the within clusters sum of squares value and the average silhouette width value are used. The within clusters sum of squares shows how close the object is to a cluster. The smaller the value, the closer the object is to a cluster. The smaller the value, the closer the object is in the cluster and how the cluster is separated from each other. Meanwhile, the average silhouette values usually range from 0 to 1, where values close to 1 indicate that the data is better grouped (Härdle & Simar, 2007).

Proverty in Indonesia

Poverty is a multidimensional issue due to its connection with the ability to access, economically, socially, cultural, political, and participation in the community. The form and factors of poverty in Indonesia are certainly influencing the policies to adress the issue. In fact, the effort of decreasing the population lives under poverty, as well as various policies and programs seen to be less effective, considering the tendency of poor people number that keep increasing over time (Nurwati, 2008).

Poverty has a lot of definitions, but it is often related to the concept of poverty from economic aspects. Various efforts to define and to identify poverty produce a concept that can be simplified. First, from a measurement point of view, poverty can be divided into absolute and relative poverty. Second, from a cause point of view, poverty can be divided into natural and structural poverty. One important condition so that alleviation poverty policies can be achieved is there must be clear criteria of who or which community group goes in the poor category and become the program target. Another condition must also be fulfilled is there must be a proper addressed about the cause of poverty in each community and area itself. Because it cannot be separated from the surrounding local values influence on the life of the people (Nurwati, 2008).

Data Set

Data used in this study was poverty indicators in Surabaya data, which were obtained from Surabaya City administration. Variable used in this study are the percentage of households with ownership status of land based on other people's property (X_1), the percentage of households with drinking water sources from meter tap (X_2), the percentage of households with 450 watts of electricity (X_3), the percentage of households with 3 kg cooking fuel gas (X_4), the percentage of households with the widest floor type other than ceramics, marble, rugs (X_5), and the percentage of households which are raskin program participants (X_6).

The steps of analysis performed in this study about poverty indicators in Surabaya were as follows:

1. All descriptive statistics were created to determine the characteristics of each village in Surabaya;
2. Cluster analysis with K-means, *Fuzzy C-Means*, *Fuzzy Gustafson Kessel, and DBSCAN* on poverty indicators in Surabaya were created;
3. The results of cluster analysis from various methods were compared with the value of *within clusters sum of square*s and the value of *average silhouette width*;
4. Conclusions are made.

Results

Statistics Description

Before creating a group analysis, it is necessary to know the missing value in all variables of the dataset, because the missing value needs to be changed first with mean value in all variables. Boxplot was used in this study to describe poverty indicators in Surabaya (see Figure 1).

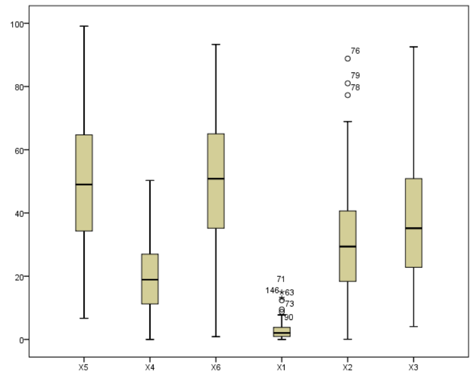

Figure 1. Boxplot Variables Proverty Indicators in Surabaya

Based on Figure 1, variable X_1, the percentage of households with ownership status of land based on other people's property, has an outlier and extreme value. While variable X_2, the percentage of households with drinking water sources from meter tap, has an outlier value. The outlier value in X_1 is located on the 90^{th}, 73^{th}, 63th, and 146th data, and the outlier value in X_2 is located on the 76th, 78th, and 79th data. In X_1, the villages with outliers are Mulyorejo (9.5 %), Kupangkrajan (8.74%), Kemayoran (12.38 %). While in X_2, the villages with outliers are Lidah Kulon (88.84 %), Lontar (77.27 %), and Made (81.02 %). The village with extreme value in X_1 is Krembangan Selatan (14.95 %). Based on the outlier value and the extreme value, percentage of households with ownership status of land based on other people's property and percentage of households with drinking water sources from meter tap was found higher in several villages than the others.

Method Implementation

In this study, the poverty villages in Surabaya were categorized based on poverty indicators using non-hierarchical methods (K-Means, Fuzzy C-Means, Fuzzy Gustafson Kessel, and DBSCAN).

K-Means

Figure 2 shows cluster plot using *K-Means* method with the number of clusters specified is 3.

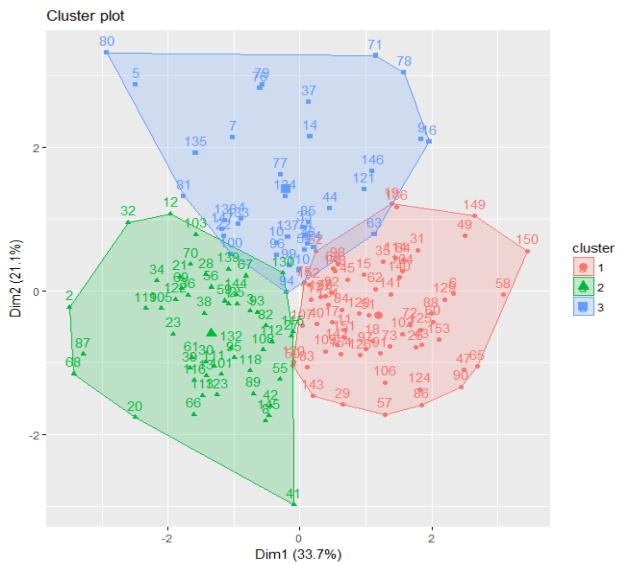

Figure 2. Plot Cluster K-Means Method for All Villages

Fuzzy C-Means

Figure 3 is a cluster plot using Fuzzy C-Means method with the number of clusters specified is 3.

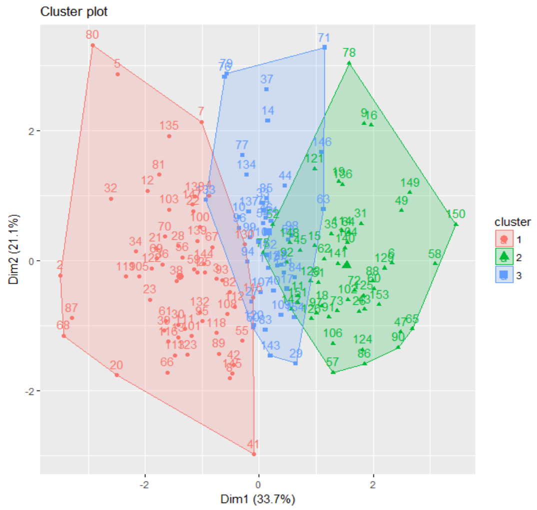

Figure 3. Cluster Plot using Fuzzy C-Means Method for All Villages

Fuzzy Gustafson Kessel

Figure 4 is a cluster plot using Fuzzy Gustafson Kessel method with the number of clusters specified is 3.

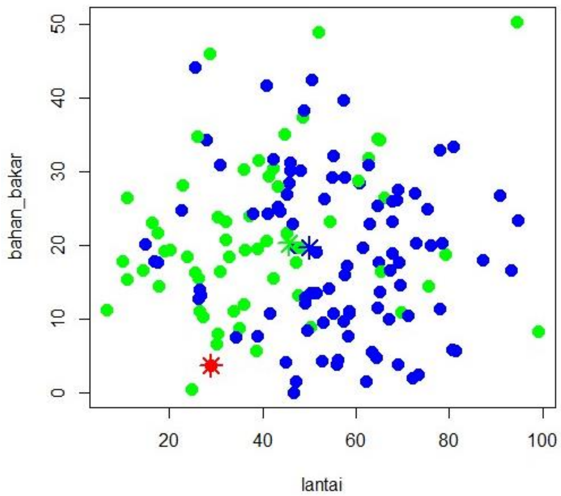

Figure 4. Cluster Plot using Fuzzy Gustafson Kessel Method for All Villages

DBSCAN

Figure 5 is cluster plot using DBSCAN method with number of clusters specified is 3.

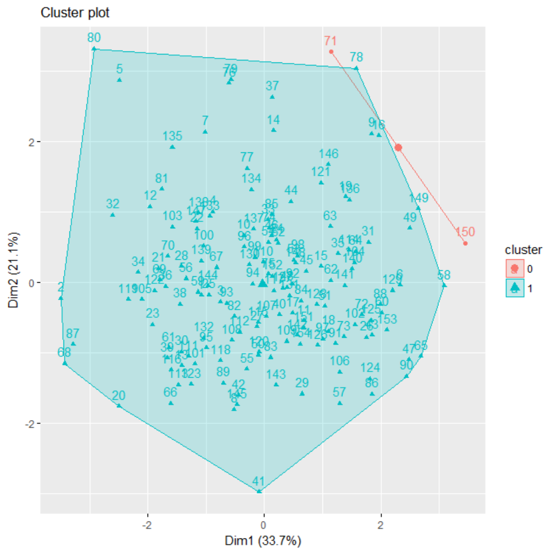

Figure 5. Cluster Plot using DBSCAN Method for All Villages

Evaluation

To get the best grouping method, the within clusters sum of squares values and the average silhouette width values obtained in each method was being compared. Table 1 shows the comparison value of each method.

Table 1. Methods Comparison

Method	Within Clusters Sum of Squares	Average Silhouette Width
K-means	**615.0695**	0.2054198
Fuzzy C-Means	640.8550	0.1756275
Fuzzy Gustafson Kessel	209288.1	0.1317502
DBSCAN	885.1028	**0.4543735**

Based on Table 1, it is found that the K-Means method was the best grouping method that was proven by the smallest value of the within clusters sum of squares compared to other methods, which was 615.0695. Meanwhile with the average silhouette width value, the DBSCAN method was found to be the best method because its value was close to 1 compared to the values of other methods.

Discussion

Based on the results analysis, it is found that by using K-Means method, the value of within clusters sums of squares was 615.0685, proving that this is the best method among others used in this study. The following research results are in accordance with (Tukiyat & Djohan, 2022), that using the sum of square within cluster (SSW) values that the K-Means algorithm is better than the DBSCAN algorithm. Furthermore, by using the value of average silhouette width, it is found that the value of K-Means method was smaller than DBSCAN, which indicates that DBSCAN method was the best method compared to the others in this study. Research results related to (Kristianto, 2021) analyzing the performance of K-Means and DBSCAN in public transportation user interest clustering DBSCAN has a good performance inclustering process. With a Silhouette Coefficient value that is greater and close to 1, based on the evaluation of these methods, K-Means and DBSCAN were the best ones. However, looking deeper, the value of within clusters sums of squares indicates the variability data is large, while the value of average silhouette width indicates clustering structure is weak. Thus, in further research researcher needs to consider reducing variability of the observation data and choosing another clustering method to increase accurate prediction.

Conclusion

Based on the analysis above, it can be concluded that there were outliers in the data and missing values in all variables. Because there were outliers value and extreme value, several villages had percentage of households with ownership status of land based on other people's property and percentage of households with drinking water sources from meter tap higher than the other villages. The village grouping in this study was based on the non-hierarchical method, namely K-mean, Fuzzy C-Means, Fuzzy Gustafson Kessel, and DBSCAN. The result of method comparison shows that K-Means was the best method for village grouping in Surabaya based on the values of within the clusters sum of squares. Meanwhile, based on the average silhouette width value, the DBSCAN method was found to be the best method because its value was close to 1 compared to the other methods. Thus, it can be implicated that K-Means and DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is the best method for village grouping in Surabaya, East Java, Indonesia in relation to poverty problems. The limitation of the research used is using 3 clusters on the K-Means and Fuzzy C-Means methods, to determine the best cluster method the Sum of Square Within value is used Cluster (SSW) and Average Silhouette.

References

Agarwal, S. (2014). Data Mining: Data Mining Concepts and Techniques. *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. https://doi: 10.1109/ICMIRA.2013.45.

Agusta, Y. (2007). K-Means-Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika*, *3*, 47–60.

Askari, S. (2021). Fuzzy C-Means Clustering Algorithm for Data with Unequal Cluster Sizes and Contaminated with Noise and Outliers: Review and Development. *Expert Systems with Applications*, *165*, 113856. https://doi.org/10.1016/j.eswa.2020.113856.

BPS. (2021). Hasil Sensus Penduduk 2020.

BPS Kota Surabaya. (2018). Profil Kemiskinan di Kota Surabaya Tahun 2017.

BPS Provinsi Jawa Timur. (2020). Kemiskinan dan Ketimpangan.

Gueorguieva, N., Valova, I., & Georgiev, G. (2017). M&MFCM: Fuzzy C-Means Clustering with Mahalanobis and Minkowski Distance Metrics. *Procedia Computer Science, 114*, 224-233. https://doi.org/10.1016/j.procs.2017.09.064.

Härdle, W., & Simar, L. (2007). Applied Multivariate Statistical Analysis: Second edition. *Applied Multivariate Statistical Analysis: Second Edition*. https://doi: 10.1007/978-3-540-72244-1.

Jalali, Z. (2016). Development of Slope Mass Rating System Using K-Means and Fuzzy C-Means Clustering Algorithms. *International Journal of Mining Science and Technology, 26*(6), 959-966. https://doi.org/10.1016/j.ijmst.2016.09.004.

Kristianto, A. (2021). Performance Analysis of K-Means and DBSCAN in Interest Clustering Use of Public Transportation. *Journal of Electronics and Computers*, Vol 14, 368-372. https://doi.org/10.51903/elkom.v14i2.551

Li, M. J., Ng, M. K., Cheung, Y. M., & Huang, J. Z. (2008). Agglomerative Fuzzy K-Means Clustering Algorithm with Selection of Number of Clusters. *IEEE Transactions on Knowledge and Data Engineering*, *20*(11), 1519–1534. https://doi: 10.1109/TKDE.2008.88.

Nurwati, N. (2008). Kemiskinan: Model Pengukuran, Permasalahan dan Alternatif Kebijakan. *Jurnal Kependudukan Padjadjaran*, *10*(1), 1–11.

Oyelade, O. J., Oladipupo, O. O., & Obagbuwa, I. C. (2010). *Application of K Means Clustering Algorithm for Prediction of Students Academic Performance*. *7*, 292–295.

Pal, N. R., & Bezdek, J. C. (1995). On Cluster Validity for the Fuzzy C-Means Model. *IEEE Transactions on Fuzzy Systems*, *3*(3), 370–379. https://doi: 10.1109/91.413225.

Park, J., Park, K. V., Yoo, S., Choi, S. O., & Han, S. W. (2020). Development of the WEEE Grouping System in South Korea Using the Hierarchical and Non-Hierarchical Clustering Algorithms. *Resources, Conservation and Recycling*, *161*, 104884. https://doi.org/10.1016/j.resconrec.2020.104884.

Raval, U. R., & Jani, C. (2016). Implementing & Improvisation of K-Means Clustering Algorithm. *International Journal of Computer Science and Mobile Computing*, *55*(5), 191–203.

Serrao, M., Chini, G., Bergantino, M., Sarnari, D., Casali, C., Conte, C., … Marinozzi, F. (2018). Identification of Specific Gait Patterns in Patients with Cerebellar Ataxia, Spastic Paraplegia, and Parkinson's Disease: A Non-

Hierarchical Cluster Analysis. *Human Movement Science*, *57*, 267–279. https://doi.org/10.1016/j.humov.2017.09.005.

Soegimo, D., & Ruswanto. (2009). *Geografi untuk SMA/MA Kelas XI*. Jakarta: Pusat Perbukuan Departemen Pendidikan Nasional.

Tilson, L. V., Excell, P. S., & Green, R. J. (1988). A Generalisation of The Fuzzy C-Means Clustering Algorithm. *Remote Sensing. Proc. IGARSS '88 Symposium, Edinburgh, 1988. Vol. 3*, *10*(2), 1783–1784. https://doi.org/10.1109/igarss.1988.569600.

Tukiyat, Djohan, Y. (2022). Analysis of the Spread of the Covid-19 Pandemic in the City of Jakarta Using the K-Means Clustering Method and Density Based Spatial Clustering of Applications With Noise. *Informatics Journal,* Vol. 9, 43-54. https://doi.org/10.31294/inf.v9i1.11226.

Treiger, B., Bondarenko, I., Van Malderen, H., & Van Grieken, R. (1995). Elucidating the Composition of Atmospheric Aerosols Through the Combined Hierarchical, Non-Hierarchical and Fuzzy Clustering of Large Electron Probe Microanalysis Data Sets. *Analytica Chimica Acta*, *317*(1), 33–51. https://doi.org/10.1016/0003-2670(95)00405-X.