applied sciences

MDPI

Article

Grouping Poverty by District/City in Central Java in 2021 Using K-Means Clustering

Muhammad Nauval Muzaki [1], Haana Lahanda Putri Samsi [2] and Muhamad Raditya Putra Marindra [3]

Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang 45363, Indonesia; muhammad21051@mail.unpad.ac.id (M.N.M.); haana21001@mail.unpad.ac.id (H.L.P.S); muhamad21037@mail.unpad.ac.id (M.R.P.M).
* Correspondence: muhammad21051@mail.unpad.ac.id

Abstract: Poverty poses a complex challenge and is a significant issue in Indonesia. The province of Central Java is characterized by a substantial population living in poverty. The aim of this research is to conduct an analysis of poverty clustering based on Districts/Cities in Central Java in 2021 using the K-Means Clustering method. As analytical material, secondary data comprising poverty indicators from the Central Statistics Agency of Central Java for the year 2021 has been utilized. The K-Means Clustering method is employed to group Districts/Cities based on similar characteristics of poverty. The research results indicate the presence of two main clusters, namely Cluster 1 with an average percentage of the population in poverty at 10.75%, and Cluster 2 with an average percentage of the population in poverty at 13.97%. Cluster 2 involves several Districts/Cities with a higher poverty rate compared to Cluster 1.

Keywords: K-Means; Clustering; No Poverty; SDGs

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

1. Introduction

In this era, poverty alleviation has become a primary focus in the global agenda. In 2015, the United Nations (UN) adopted the 2030 Agenda for Sustainable Development, encompassing 17 Sustainable Development Goals (SDGs) as a guide for the global community towards a fairer, sustainable, and inclusive world. One of the pivotal and highly relevant SDGs in this context is Sustainable Development Goal number 1: "No Poverty.

According to the data from the Sustainable Development Report 2020 [1], Indonesia has progressed from the previous year with a score of 64.2 to 65.3, succeeding in raising 4 goals, one of which is to eradicate poverty, 6 goals have improved successfully, while 7 goals are stagnant and even decreased, one of which is reducing inequality (Sachs et al, 2020)

This goal is set to eradicate all forms and dimensions of poverty by 2030. Poverty is not only related to the availability of material resources but also includes aspects such as access to education, healthcare services, decent employment, and participation in decision-making processes. Poverty alleviation serves as an essential foundation for achieving other sustainable development goals, as it often constitutes the root of global issues such as inequality, hunger, and social instability.

This section will discuss several key aspects of Sustainable Development Goal number 1, providing an overview of the multidimensional complexity of poverty and detailing its relevance and impact on global society. Additionally, it will outline

strategies and initiatives that have been implemented or are still needed to achieve this ambitious target.

One way to predict regional groups based on poverty criteria is through the application of the K-Means Clustering method. This method groups n objects into different classes based on their distance from the cluster center. Cluster analysis, as an analytical tool, is useful for summarizing data. The process of simplifying data can be done by grouping objects based on certain characteristics that are similar among the objects being studied.

The attributes used in this study are the percentage of poor population (%), poverty depth index (%), poverty severity index (%), human development index, gender development index, distribution of percentage of households with access to adequate drinking water sources (%), population aged 15 and over who are not in the labor force (million), and percentage of household by district/city and main fuelwood used for (%).

Based on an understanding of the complex issues surrounding poverty, it is hoped that this research can make a positive contribution to collective efforts in achieving Sustainable Development Goal number 1 and supporting the transformation of the world towards a more sustainable, fair, and inclusive life for all.

2. Literature Review

Sustainable Development Goals

Sustainable development is defined as a social movement undertaken by a group of individuals with similar ideologies, aiming to achieve desired goals in meeting present needs while concurrently preserving the integrity of Earth's life-support systems. It serves as the foundation for the well-being of both the current and future generations, ultimately guiding the world towards a sustainable direction [2]. Sustainable development is construed as a form of social development, wherein it fulfills the needs of the current generation without compromising the ability of future generations to meet their own needs towards prosperity [3,4].

Poverty

In Indonesia, poverty remains a serious issue, particularly in connection with the widening gap between the rich and the poor. As a member of the United Nations, Indonesia is committed to addressing this problem in line with the Sustainable Development Goals (SDGs) declaration. This means that Indonesia is also required to achieve the targets set in the UN declaration. The government's efforts to comprehensively address poverty date back to 1995 with the issuance of the Inpres Desa Tertinggal. Through Presidential Regulation No. 15 of 2010 on Accelerating Poverty Alleviation, the government established the National Team for Accelerating Poverty Alleviation (TNP2K), directly chaired by the Vice President. This national effort indicates that poverty remains a serious issue. The central government has even realized the disbursement of the first stage of village funds, approximately 47 trillion rupiahs, to village governments through the Ministry of Finance. The utilization of these village funds is then monitored by the Ministry of Village, Development of Disadvantaged Regions, and Transmigration (Kemendesa PDTT) to ensure compliance with the established Ministerial Regulations. According to Ministerial Regulation of Village, Development of Disadvantaged Regions, and Transmigration No. 21 of 2015 concerning the Determination of Priority Uses of Village Funds, the village funds for the year 2016 were allocated to finance the implementation of local-scale programs and activities in the fields of Village Development and Empowerment of Village Communities.

Cluster Analysis

Clustering is the process of grouping similar objects into different clusters, or more precisely, partitioning a data set into subsets so that data within each subset has meaningful significance. A cluster consists of a collection of objects that are similar to each other and different from those in other clusters. Clustering algorithms are divided into two parts: hierarchical and partitional. Hierarchical algorithms find clusters sequentially where clusters are predetermined, while partitional algorithms determine all groups at a specific time [5]. Clustering can also be described as a process that groups and divides data patterns into several datasets, forming similar patterns grouped in the same cluster and separating distinct patterns into different clusters [6].

Clustering, also known as unsupervised learning, involves algorithms that can identify distinctive structures in similarities or differences, such as distances, among individual data points in a dataset [7]. The definition of clustering is the process of organizing data into groups with high similarity within a group, while data with dissimilarity is placed in different groups [8]. The degree of proximity or similarity of data patterns is measured using various predefined distance metrics. The measurement of similarity between two objects is done using a distance measure [9]. There are many distance metrics proposed for data clustering [10], and one commonly used distance metric for measuring similarity between data objects is Euclidean (Euclidean Distance) [11].

$$d_E = d_{i,j} = ||x_j - v_i||^2$$

Non-hierarchy Method

This approach begins by forming a desired number of clusters as an initial step. Once the number of clusters is determined, several grouping processes are carried out without following a specific hierarchical process. In techniques like K-means, objects can switch between clusters at each grouping stage.

The steps for placing observation objects in the K-means algorithm are as follows:

1. Determine the number of groups.
2. Determine the initial K-centers of the groups.
3. Calculate the distance of each object to each group center.
4. If the distance from object i to group center j is the smallest among the group centers, then object i becomes a member of group j. (i = 1,2,...,n ; j = 1, 2, … , k.)
5. Redetermine the group centers obtained from the average values of its group members.
6. If the formed groups do not change or there are no observation objects that move groups, the iteration stops. If not, go back to step 2.

K-Means Clustering

In the paper published by Elsevier "Data clustering: 50 years after K-means" [11], it was revealed that organizing data into clusters is the most basic model for understanding and learning. Cluster analysis is the formal study of grouping or classifying objects according to characteristics measured by their similarity to each other. Clustering involves grouping using unsupervised learning techniques, does not require a training phase, and does not use labels within each group.

Clustering methods divide data into groups in such a way that data with similar characteristics are grouped into the same cluster. The goal of clustering is to minimize

the objective function specified in the clustering process, which usually seeks to minimize variation within clusters and maximize variation between clusters.

Clustering or Clustering analysis is the process of forming groups of data (clusters) from a set of data, unknown to the group or class, and the process of determining which cluster the data belongs to. Clustering is a process for identifying taxonomic or geological classes or analyzing the topology of existing data. From a data mining perspective, clustering is not a clarification process. Because during the clarification process, the data is grouped into pre-known classes.

3. Material and Method

Data Source

The dataset used in this research consists of secondary data extracted from the 2021 publication of Central Java Province by the Central Statistics Agency (Badan Pusat Statistik Jawa Tengah) [12]. This dataset comprises eight variables and includes information on Social and People's Welfare in Central Java Province for each district/city. The type of data used in this research is ratio data.

Research Variables

The data variables obtained from the 2021 Central Java Province BPS website, used in this research, total eight variables. The following are the variables used in the study:

Variabel	Indikator
X1	Percentage of Poor Population (%)
X2	Poverty Depth Index (P1) (%)
X3	Poverty Severity Index (P2) (%)
X4	Human Development Index
X5	Gender Development Index
X6	Distribution of Percentage of Households with Access to Adequate Drinking Water Sources (%)
X7	Population Aged 15 and Over Who Are Not in the Labor Force (Million)
X8	Percentage of Household by District/City and Main Fuelwood Used for (%)

Calculating Euclidean Distance

The Euclidean distance measurement method is the simplest and most commonly used distance calculation in research. The Euclidean distance equation can be expressed in the following form, with dij representing the distance between object i and object j, and k being the variable of (1,2,3,...p), with the condition $i \neq j$.

$$d_{i,j} = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2}$$

Determining the Number of Clusters

Establishing the number of clusters is a crucial aspect in cluster analysis. The determination of the number of clusters is based on the consideration that having too few groups may result in significant variation within a large group, while having too many groups may introduce unwanted complexity. Some guidelines used to determine the number of clusters involve several methods, one of which is the silhouette method.

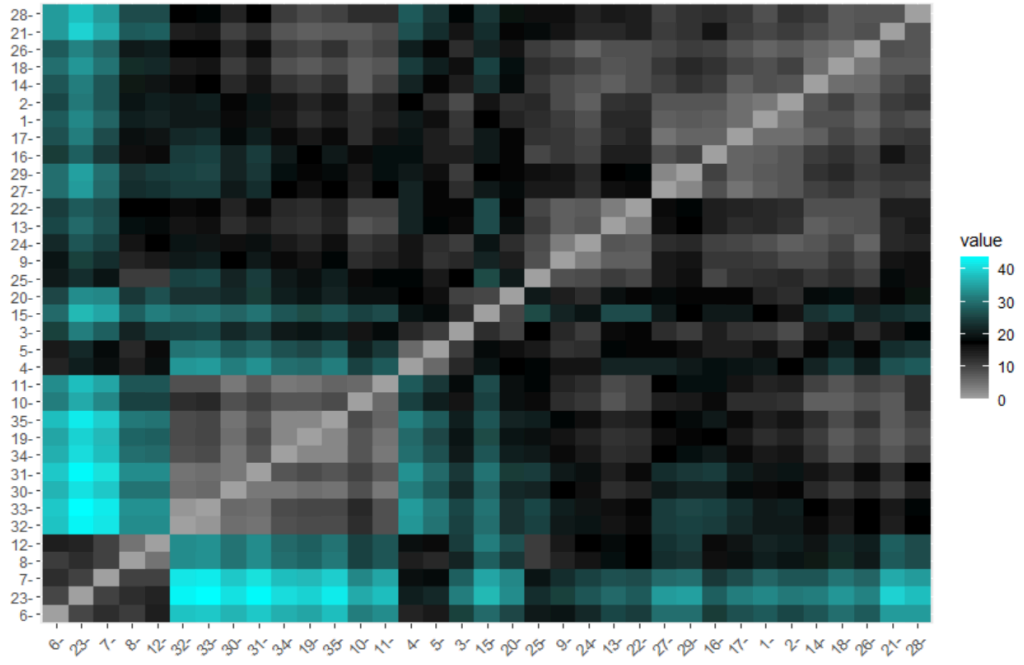
Interpreting Formed Clusters

Interpreting clusters involves evaluating the centroids, which are the average values of each variable for the objects within that cluster. The use of centroid values facilitates the labeling or naming of each cluster, and the determination of cluster labels may also depend on the goals or benefits desired.

4. Result and Discussion

In this chapter, the research results will be explained, divided into several subsections covering Euclidean distance visualization, determining the number of formed clusters, and the interpretation of the K-Means Cluster analysis.

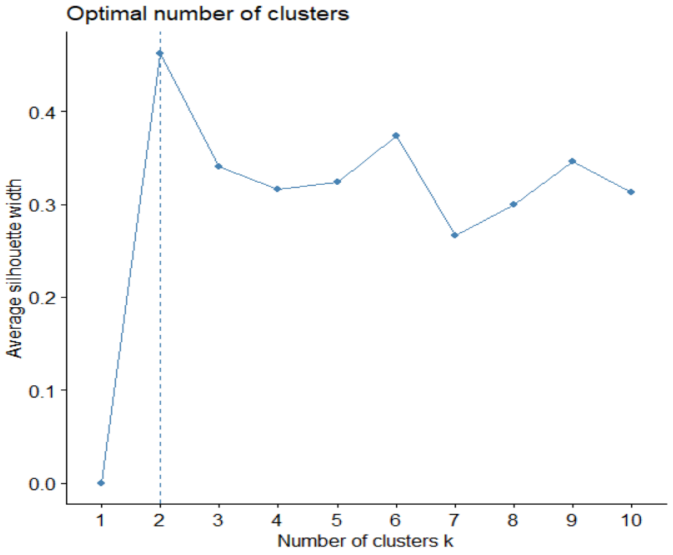
Visualizing Euclidean Distance



From the above Euclidean distance visualization, a distance matrix is generated with a color gradient ranging from bright to darker shades, eventually reaching a light blue color. This visualization is useful for assessing patterns or structures within the distance matrix, where color variations reflect the level of proximity or distance between data points.

Determining the Number of Clusters

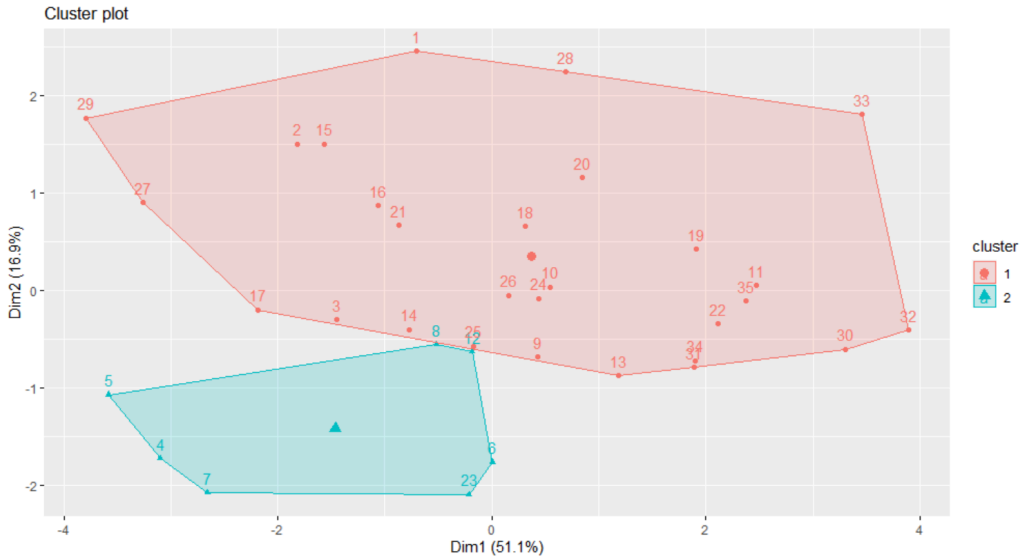
Using the silhouette method, we can visually evaluate the average Silhouette values for various possible numbers of clusters, aiding in determining the number of clusters that yield the best results in the context of cluster analysis using K-means.



Based on the results of determining the number of clusters using the silhouette method, the ideal number of clusters obtained is 2 clusters. This means that the classification of poverty in the regions of Central Java Province in 2021 will be grouped into 2 clusters.

Interpreting K-Means Cluster

To view the districts/cities in Central Java Province in each cluster, data visualization is performed by creating a cluster plot as follows.



From the results of the above image, it can be observed the members of the two clusters consisting of districts/cities in Central Java Province in the year 2021, classified as follows.

Cluster 1

No	Cluster 1	No	Cluster 1
1	Kabupaten Cilacap	15	Kabupaten Demak
2	Kabupaten Banyumas	16	Kabupaten Temanggung
3	Kabupaten Purbalingga	17	Kabupaten Kendal
4	Kabupaten Boyolali	18	Kabupaten Batang

5	Kabupaten Klaten	19	Kabupaten Pekalongan
6	Kabupaten Sukoharjo	20	Kabupaten Pemalang
7	Kabupaten Karanganyar	21	Kabupaten Tegal
8	Kabupaten Sragen	22	Kabupaten Brebes
9	Kabupaten Grobogan	23	Kota Magelang
10	Kabupaten Blora	24	Kota Surakarta
11	Kabupaten Rembang	25	Kota Salatiga
12	Kabupaten Pati	26	Kota Semarang
13	Kabupaten Kudus	27	Kota Pekalongan
14	Kabupaten Jepara	28	Kota Tegal

Cluster 2

No	Cluster 2
1	Kabupaten Banjarnegara
2	Kabupaten Kebumen
3	Kabupaten Purworejo
4	Kabupaten Wonosobo
5	Kabupaten Magelang
6	Kabupaten Wonogiri
7	Kabupaten Semarang

Table of Average Percentage of Poor Population

Average Percentage of Poor Population	
Cluster 1	10.75%
Cluster 2	13.97%

Based on the average percentage of poor population in Districts/Cities in Central Java Province in 2021, it is found that the highest average is in cluster 2 with an average percentage of poor population of 13.97%. Meanwhile, cluster 1 has an average percentage of poor population of 10.75%.

5. Conclusion

Overall, the poverty situation in Central Java can be observed through several indicators, including the number of people living in poverty, the depth and severity of poverty, the highest human development index, the highest gender development index, families that have not achieved prosperity, inadequate sanitation conditions, unsafe drinking water sources, unemployed populations, and the use of cooking fuels such as firewood, charcoal, or kerosene.

The conclusion drawn from the application of non-hierarchical cluster analysis method with the aim of classifying a group of Kabupaten/Kota in Jawa Tengah Province in 2021, totaling 35 observations, into several clusters based on similarities or common characteristics among observations, resulting in observations within the same cluster having greater similarity than objects outside the cluster. The outcome is that by determining 2 clusters using the silhouette method and K-means cluster analysis as a tool to group them, the cluster groups and members are as follows.

The cluster with the highest poverty rate is cluster 2, consisting of Banjarnegara District, Kebumen District, Purworejo District, Wonosobo District, Magelang District, Wonogiri District, and Semarang District with a percentage of poor population at 13.97%.

This is followed by Cluster 1, consisting of Kabupaten Cilacap, Kabupaten Banyumas, Kabupaten Purbalingga, Kabupaten Boyolali, Kabupaten Klaten, Kabupaten Sukoharjo, Kabupaten Karanganyar, Kabupaten Sragen, Kabupaten Grobogan, Kabupaten Blora, Kabupaten Rembang, Kabupaten Pati, Kabupaten Kudus, Kabupaten Jepara, Kabupaten Demak, Kabupaten Temanggung, Kabupaten Kendal, Kabupaten Batang, Kabupaten Pekalongan, Kabupaten Pemalang, Kabupaten Tegal, Kabupaten Brebes, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan, dan Kota Tegal with a percentage of poor population at 10.75%.

Future research potential could focus on specific aspects that may require further attention. For example, in-depth research could be conducted to explore factors contributing to the depth or severity of poverty. Additionally, further investigation into the impact of implemented welfare policies or programs could provide a deeper understanding of the effectiveness of these measures in addressing poverty.

By continually deepening the understanding of the characteristics and dynamics of poverty, it is hoped that this research can make a meaningful contribution to designing more effective policies to reduce and alleviate poverty in Districts/Cities in Central Java Province. Furthermore, it is hoped that this research can inspire future studies focusing on this social issue context.

References

1. Sachs, J., Schmidt-Traub, G., Kroll, C., Lafortune, G., Fuller, G., Woelm, F. 2020. The Sustainable Development Goals and COVID-19. Sustainable Development Report 2020. Cambridge: Cambridge University Press.
2. Erlinda, N. (2016). Kebijakan pembangunan wilayah berkelanjutan di provinsi jambi melalui pendekatan model flag. Jurnal Ekonomi Dan Kebijakan Publik, 7(1), 1. https://doi.org/10.22212/jekp.v7i1.416
3. Jayanti, E. (2017). Hubungan Pertumbuhan Penduduk Dengan Tujuan Pembangunan Berkelanjutan Di Sumatera. Jurnal Ekonomi Dan Kebijakan Publik Indonesia, 4(2), 136– 155.
4. Nababan, Y. J., Syaukat, Y., Juanda, B., & Sutomo, S. (2014). TANTANGAN BAGI PEMBANGUNAN BERKELANJUTAN DI KALIMANTAN TIMUR: MENUJU INCLUSIVE GREEN ECONOMY. Masyarakat Indonesia, 40(2), 211–228.
5. Madhulatha, T.S., 2012. An Overview On Clustering Methods. IOSR Journal of Engineering, II(4), pp.719-725
6. HUNG, C.M., WU, J., CHANG, J.H. & YANG, D.L., 2005. An Efficient k-Means Clustering Algorithm Using Simple Partitioning. JOURNAL OF INFORMATION SCIENCE AND ENGINEERING, XXI(1), pp.1157-77
7. Cios, Krzysztof J., et al. 2007. Data Mining : A Knowledge Discovey Approach. New York : Springer Science+Business Media, LLC.
8. Harrington, P. 2012. Machine Learning in Action. Manning Publications Co, Shelter Island, New York.
9. Xu, R. and Wunsch, D.I.I. (2005) Survey of Clustering Algorithms. IEEE Transactions on Neural Networks, 16, 645- 678.
10. Pandit, S., & Gupta, S. (2011). A Comparative Study on Distance Measuring Approaches for Clustering. International Journal of Research in Computer Science, 2(1), 29–31.
11. Jain. Data clustering: 50 years beyond K-means. 2009
12. Badan Pusat Statistik Jawa Tengah, 2021.
13. F. Abascal and A. Valencia, “Clustering of proximal sequence space for the identification of protein families,” Bioinformatics, vol. 18, pp. 908–921, 2002.
14. Zenklinov AP. 2021. Pengelompokkan Kemiskinan Kabupaten/Kota di Jawa Tengah Meggunakan Algoritma Fuzzy Possibilistic C-Means dan Possibilistic Fuzzy C-Means.
15. Aggarwal, C. C., & Zhai, C. (2012). Mining Text Data (C. C. Aggarwal & C. Zhai (eds.)). Springer. https://doi.org/10.1007/978-1 - 4614-3223-4).
16. Gorunescu, F. (2011). Data Mining: Concepts, Models, and Techniques (J. Kacprzyk & L. C. Jain (eds.); 12th ed.). Springer. https://doi.org/10.1007/978-3-642-19721 -5.
17. Hudaya, D. (2009). Faktor-Faktor yang Mempengaruhi Tingkat Kemiskinan di Indonesia. Institut Pertanian Bogor.
18. Ishartono, & Raharjo, S. T. (2015). Sustainable Development Goals (SDGs) dan Pengentasan Kemiskinan. Social Work Jurnal, 6(2), 154– 272.

19. Mann, A. K., & Kaur, N. (2013). Survey Paper on Clustering Techniques. International Journal of Science, Engineering and Technology Research (IJSTER), 2(4), 803–806.
20. Suryawati, C. (2005). Memahami kemiskinan secara multidimensional. Jurnal Manajemen Pelayanan Kesehatan (JMPK), 08(03), 121 – 129.
21. E. Muningsih1 dan S. Kiswati. —Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang‖, Jurnal Bianglala Informatika, Vol. 3, No. 1, 2015.
22. M. Anggara, H. Sujaini, H. Nasution, "Pemilihan Distance Measure Pada K-Means Clustering untuk Pengelompokkan Member DiAlvaro Fitness", Jurnal Sistem dan Teknologi Informasi (JUSTIN), Vol. 4, No. 1, 2016.
23. Suryawati, C. (2005). Memahami kemiskinan secara multidimensional. Jurnal Manajemen Pelayanan Kesehatan (JMPK), 08(03), 121 – 129.
24. Hudaya, D. (2009). Faktor-Faktor yang Mempengaruhi Tingkat Kemiskinan di Indonesia. Institut Pertanian Bogor.
25. Ishartono, & Raharjo, S. T. (2015). Sustainable Development Goals (SDGs) dan Pengentasan Kemiskinan. Social Work Jurnal, 6(2), 154–272.
26. Madhulatha, T.S., 2012. An Overview On Clustering Methods. IOSR Journal of Engineering, II(4), pp.719-725
27. HUNG, C.M., WU, J., CHANG, J.H. & YANG, D.L., 2005. An Efficient k-Means Clustering Algorithm Using Simple Partitioning. JOURNAL OF INFORMATION SCIENCE AND ENGINEERING, XXI(1), pp.1157-77.
28. http://www.bappenas.go.id/id/berita-dan-siaranpers/sektor-infrastruktur-prioritas-penggunaan-danadesa-2016/