Vol. 2 No. (2) (December 2024) 72-79

Journal of Data Insights

http://journalnew.unimus.ac.id/index.php/jodi

Fuzzy Gustafson Kessel for Infrastructure Development Strategy in South Sumatra Province

Oktaviana Rahma Dhani[1], Febi Anggun Lestari[2], Zahra Aura Hisani[3], Ariska Fitriyana Ningrum[4*], Alwan Fadlurohman

[1]Universitas Muhammadiyah Semarang, Indonesia
[2]Universitas Muhammadiyah Semarang, Indonesia
[3]Universitas Muhammadiyah Semarang, Indonesia
[4]Universitas Muhammadiyah Semarang, Indonesia
[5]Universitas Muhammadiyah Semarang, Indonesia

DOI: https://doi.org/10.26714/jodi.v2i2.650

Article Information

Article History:
Submitted: 10th December 2024
Revised: 16th December 2024
Accepted: 31th December 2024

Keywords:
Infrastructure Development; Clustering; Fuzzy Gustafson Kessel; Data Mining.

Abstract

Infrastructure development is a strategic element in improving public services and economic growth. South Sumatra Province, with its large economic potential, faces challenges in managing efficient and sustainable infrastructure development. This research aims to apply the Fuzzy Gustafson Kessel (FGK) method in decision making related to infrastructure development in South Sumatra Province. FGK combines fuzzy logic with Gustafson Kessel clustering algorithm to handle uncertainty and data variation from various stakeholders. The data used in this study includes population and geographic census data from the Central Bureau of Statistics of South Sumatra Province in 2023, with five indicators: population, area, population growth rate, population density, and poverty rate. The results show that South Sumatra is divided into three main clusters based on its infrastructure and demographic characteristics. This clustering is expected to improve the effectiveness and efficiency of infrastructure development decision-making, provide more appropriate policy recommendations, and potentially be applied in other regions with similar challenges.

✉ Corresponding Author:
E-mail: ariskafitriyana@unimus.ac.id

e-ISSN: 2988 - 2109

INTRODUCTION

Infrastructure development has a very strategic role because it is closely related to public services. With the availability of good and adequate infrastructure, transportation modes become smooth and can ultimately improve the economy, people's welfare, dignity, and regional competitiveness at the global level (Wahyu Abadi, Guntoro, & Prajarto, 2014). In general, the definition of infrastructure can be explained as a system of physical facilities that support the life, sustainability, and economic and social growth of a society or community (Shobirin, & Ali, 2019).

South Sumatra Province is one of the important provinces in Indonesia with great economic potential (BPS South Sumatra Prov, 2023). South Sumatra has an area of about 86,772 square kilometers with a growing population, reaching 8.7 million people in 2022. The province is rich in natural resources from various sectors such as the mining and energy sector, the agricultural sector, the processing industry sector, and the trade sector, which makes South Sumatra one of the significant economic centers in Indonesia. However, the challenge of efficient and sustainable infrastructure development remains a major focus in an effort to improve economic growth and community welfare. The central and provincial governments have allocated significant budgets for infrastructure development in South Sumatra. This is reflected in government projects regarding public infrastructure development (BPS Prov South Sumatra, 2023).

While these efforts are geared towards improving connectivity and increasing accessibility in the region, infrastructure-related decision-making is often faced with data complexity as well as diverse preferences from various stakeholders. Inappropriate decisions can negatively impact project efficiency and long-term benefits. In an effort to address the complexity of infrastructure-related decision-making faced with uncertainty and data variations from multiple stakeholders, innovative approaches such as Fuzzy Gustafson Kessel (FGK) offer an attractive solution (Atakishiyev, S & Reformat, M. Z., 2021). The FGK method integrates Fuzzy Logic with the Gustafson Kessel clustering algorithm, which enables more comprehensive and accurate analysis in the face of uncertainty and data variation (Dey, Subhrasankha & Tanmoy Dam, 2021).

By combining the advantages of Fuzzy Logic in handling ambiguity and the Gustafson Kessel clustering algorithm that is able to detect data groups with elliptical shapes, FGK can provide more accurate clustering results and is robust to noise or outliers in the data. Previous research shows that the effectiveness of FGK in classifying multidimensional data with complex and elliptical cluster shapes, as well as in data conditions with high noise, is superior to other fuzzy clustering methods, especially in terms of accuracy and robustness to noise (Hosein Hashemi, Abdolrahim Javaherian, & Rob, 2008).

This research aims to explore and apply the Fuzzy Gustafson Kessel (FGK) method in the context of infrastructure development decision making in South Sumatra Province. Specifically, this research analyzes the challenges and complexity of data in infrastructure decision making, implements the FGK model to cluster infrastructure project data, evaluates the performance of FGK in clustering data, assesses the contribution of FGK to the quality of decision making, and provides policy recommendations to improve the effectiveness and efficiency of infrastructure development decisions based on the research results. In addition, the findings of this study have the potential to be applied by other regions in Indonesia that face similar challenges in infrastructure development.

LITERATURE REVIEW

A. DATA

Data is a collection consisting of facts to provide a broad picture related to a situation. Someone who will make a policy or decision will generally use data as a consideration. Through data one can analyze, describe, or explain a situation.

B. DATA MINING

Data mining is the process of extracting / extracting / finding hidden information from large databases and is useful for interested parties. Data mining can help data holders to analyze and find unexpected relationships between data, which in turn will help decision making (Kurniawan, Sarjon, & Sumijan, 2020).

Data mining can be said to be the process of extracting from a large amount of available data (Pramadhani & Setiadi, 2014). The knowledge gained from the data mining process includes valuable insights into hidden patterns, trends, and relationships in large and complex data, which can support strategic and operational decision making in various fields.

C. CLUSTERING

Clustering is a data mining technique used to analyze data to solve problems in grouping data or rather partitioning a dataset into subsets. In clustering techniques, the target is to distribute cases (objects, people, events and others) into a group, so that the degree of connection between members of the same cluster is strong and weak between members of different clusters (Wardhani, 2016).

D. FUZZY GUSTAFSON KESSEL

Fuzzy Gustafsson-Kessel (FGK) is a way of grouping data where the existence of each data in a group is determined by the membership value and this technique is a variant of fuzzy clustering developed by Gustafson and Kessel in 1979 (Mauliyadi M, Hizir Sofyan, & Muhammad Subianto, 2015).

The algorithm of Gustafson-Kessel clustering is as follows:

$$J(X;U,V,A) = \sum_{i=1}^{n} \sum_{k=1}^{c} (\mu_{ik})^{W} \, (d_{ikAi})^2$$

With

Where :

X is the data to be grouped

$$d_{ikAi} = \left(X_{ij} - X_{kj}\right)^{T} A_i \left(X_{ij} - V_{kj}\right)$$

1. U is the initial partition matrix by generating random values
2. V is the group center matrix
3. Ai is indicating the group distance
4. The steps in the FGK method are as follows:
5. Determine the data to be grouped in the form of an X matrix of size n x m, where n = the number of data to be grouped, m = the attributes of each data.
6. Determine:
7. a. The number of groups to be formed = c (≥ 2).
8. b. Rank or weight w = 2.
9. c. Maximum iterations.
10. d. Least stopping criterion = ξ
11. e. Initial iteration t = 1.
12. Form the initial partition matrix U by generating random values μik, i=1,2,...,n; k=1,2,...,c; as the elements of the initial partition matrix U as follows:

$$U = \begin{bmatrix} \mu_{11}(x_1) & \dots & \mu_{1n}(x_n) \\ \dots & \dots & \dots \\ \mu_{c1}(x_1) & \dots & \mu_{cn}(x_n) \end{bmatrix}$$

13. Calculate the group center v_i, for v each group as follows:

$$V_{kj} = \frac{\sum_{i=1}^{n}\left((\mu_{ik})^W X_{ij}\right)}{\sum_{i=1}^{n}(\mu_{ik})^W}$$

14. Calculate the clustering matrix covariance (F_i) with the formula

$$F_i = \frac{\sum_{i=1}^{n}(\mu_{ik})^W\left(X_{ij} - V_{kj}\right)^T}{\sum_{i=1}^{n}(\mu_{ik})^W}$$

15. Calculating the distance

$$d_{ikAi} = \left(X_{ij} - V_{kj}\right)^T \left[(\rho_i \det(F_I)^{\frac{1}{n}} F_i{}^{-1}\right] \left(X_{ij} - V_{kj}\right)$$

16. Fixing the membership degree of each data in each group with equation :

$$\mu_{ik} = \left[\sum_{j=1}^{k} \left(\frac{d_{ikAi}}{d_{jkAi}}\right)^{\frac{2}{(W-1)}}\right]^{-1}$$

17. Determine the stopping criteria If (|Pt-Pt-1|≤ ε or t>MaxIter) then it is stopped, but if not. Then the element is increased by iteration t=t+1 and return to step 4.

METHODS

A. RESEARCH DATA

The data used for this research are population census data and geographic data. The data taken is data from the Central Bureau of Statistics (BPS) of South Sumatra Province in the form of South Sumatra data in numbers in 2023. The observation object consists of 17 districts / cities using 5 indicators. The following indicators are used:

1. Total Population (X1);
2. Area (X2);
3. Population Growth Rate (X3);
4. Population Density (X4);
5. Poverty Rate (X5).

B. RESEARCH STEPS

The research steps for grouping districts / cities in South Sumatra Province using Fuzzy Gustafson Kessel are as follows:

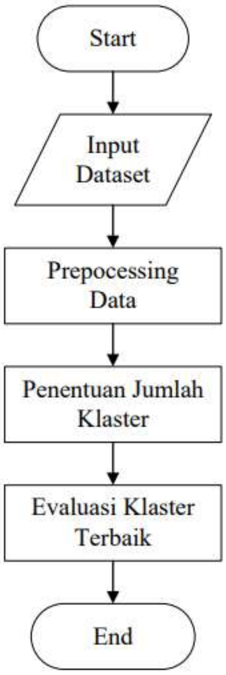
1. Data Collection

   The data used is secondary data obtained from the Central Bureau of Statistics (BPS) of South Sumatra Province in figures for 2023.

2. Data Preprocessing

   At this stage, data entry is carried out, looking at the characteristics of the data, then converting the data into numeric. Furthermore, the data normalization process is carried out. The purpose of data normalization itself is to change the value of the numeric column in the data set to use a common scale, without distorting differences in the range of values or losing information.

3. Determination of the Number of Clusters
   Determination of the best number of clusters is done by finding the number of clusters that provide the highest silhouette value, because a larger silhouette value indicates that the objects in the cluster are more similar to each other than the objects in other clusters.
4. Best Cluster Evaluation Grouping of district/cities using Fuzzy Gustafson Kessel based on the best group



Picture 1. Research Steps

RESULT AND DISCUSSION

This research uses a dataset from the website of the Central Bureau of Statistics of South Sumatra Province in 2023 with variables X_1 is Population, X_2 is Regional Area, X_3 is Population Growth Rate, X_4 is Population Density, and X_5 Poverty Level. Based on the results of the shilhoutte score, the optimal K value is 3. So that the cluster to be searched becomes three parts, namely cluster C1 = low population with low area, cluster C2 = with average population and average area, and cluster C3 with a dense population with a large area. The results of cluster evaluation using fuzzy sillhoutte are as follows.

Table 1. Result of Cluster Evaluation

Cluster	Validation Score
2	0.04887266
3	0.15311595
4	-0.20340262
5	-0.23750431
6	-0.11521775

Based on the results of the best cluster evaluation, the highest validity value is 0.15 with a total of 3 clusters. Furthermore, clustering is carried out using fuzzy Gusftson Kessel. Grouping Kab / City in South Sumatra Province using FGK obtained results in the first cluster consisting of Kab / City of Ogan Komering Ulu, Ogan Komering Ilir, Muara Enim, Musi Rawas, Musi Banyuasin, and Banyuasin. While the second cluster consists of Lahat, South Ogan Komering Ulu, Ogan Ilir, Pali, North Musi Rawas, prabumulih, and lubuk linggau while the districts/cities that are included in the 3rd cluster are east ogan komering ulu, empat lawang, palembang, and pagar alam.

.

Picture 2. Map of Regency/City Clustering in South Sumatra Province

Table 2 Characteristics of each cluster

Indikator	Cluster 1	Cluster 2	Cluster 3
Total Population (X_1)	622771.33	306067.71	716105
Area (X_2)	10090.98	2663.43	1895.44
Population Growth Rate (X_3)	1.31	1.37	0.96
Population Density (X_4)	70.51	238.07	1345.71
Poverty Level (X_5)	12.36	13.09	10.22

Based on Table 2, the characteristics of each cluster show substantial variation among the three clusters. Cluster 1 has an average value on variable X_1 (Total Population) that is medium, X_2 (Area Size) that is high, X_3 (Population Growth Rate) that is medium, X_4 (Population Density) that is low, and X_5 (Poverty Rate) that is medium, indicating that this region has a large area with a medium population and poverty rate and low population density. Cluster 2 has a low average value for the variable X_1 (Total Population), a medium X_2 (Total Area), a high X_3 (Population Growth Rate), a medium X_4 (Population Density), and a high X_5 (Poverty Level), indicating that this region has rapid population growth and a high poverty rate despite its low population. Cluster 3 has a high average value for the variable X_1 (Population Size), a low X_2 (Area Size), a low X_3 (Population Growth Rate), a high X_4 (Population Density), and a low X_5 (Poverty Rate), indicating that this region has a high population size and density, but with a low population growth rate and poverty rate.

CONCLUTION

Based on the results of clustering using Fuzzy Gustafson Kessel on infrastructure development strategies in South Sumatra Province by looking at variables X_1 (Total Population), X_2 (Area), X_3 (Population Growth Rate), X_4 (Population Density), and X_5 (Poverty Level) it can be concluded that the South Sumatra region is divided into three main clusters. Cluster 1 includes regions with large areas, moderate population and poverty rates, and low population density. The regions included in cluster 1 are Ogan Komering Ulu, Ogan Komering Ilir, Muara Enim, Musi Rawas, Musi Banyuasin, and Banyuasin. Cluster 2 includes regions with low population, fast population growth, and high poverty rates. The regions included in cluster 2 are Lahat, South Ogan Komering

Ulu, Ogan Ilir, Pali, North Musi Rawas, Prabumulih, and Lubuk Linggau. Cluster 3 includes regions with high population size and density, but low population growth rates and poverty rates. The regions included in cluster 3 are East Ogan Komering Ulu, Empat Lawang, Palembang, and Pagar Alam. This clustering provides a strong basis for more informed and efficient decision-making in infrastructure development in South Sumatra Province. In particular, cluster 3, which includes Palembang and its surroundings, requires more attention in infrastructure development given the high number and density of population in the area. By focusing on infrastructure development in cluster 3, it is expected to increase economic growth and community welfare in a sustainable manner.Based on the clustering results using Fuzzy Gustafson Kessel on infrastructure development strategies in South Sumatra Province by looking at variables X_1 (Total Population), X_2 (Area Size), X_3 (Population Growth Rate), X_4 (Population Density), and X_5 (Poverty Level) it can be concluded that the South Sumatra region is divided into three main clusters. Cluster 1 includes regions with large areas, moderate population and poverty rates, and low population density. The regions included in cluster 1 are Ogan Komering Ulu, Ogan Komering Ilir, Muara Enim, Musi Rawas, Musi Banyuasin, and Banyuasin. Cluster 2 includes regions with low population, fast population growth, and high poverty rates. The regions included in cluster 2 are Lahat, South Ogan Komering Ulu, Ogan Ilir, Pali, North Musi Rawas, Prabumulih, and Lubuk Linggau. Cluster 3 includes regions with high population size and density, but low population growth rates and poverty rates. The regions included in cluster 3 are East Ogan Komering Ulu, Empat Lawang, Palembang, and Pagar Alam. This clustering provides a strong basis for more informed and efficient decision-making in infrastructure development in South Sumatra Province. In particular, cluster 3, which includes Palembang and its surroundings, requires more attention in infrastructure development given the high number and density of population in the area. By focusing on infrastructure development in cluster 3, it is expected to improve economic growth and community welfare in a sustainable manner.

REFERENCE

Totok Wahyu Abadi, Nunung Prajarto, Budi Guntoro, "Performance E-Government Untuk Peningkatan Partisipasi Masyarakat Dalam Pembangunan Infrastruktur Di Kabupaten Sidoarjo,"Vol.4,No.3,225-330,2014.

Muhammad Shobirin, Hapzi Ali,"Strategi Pengembangan Infrastruktur Dalam Meningkatkan Pelayanan Penumpang Di Bandar Udara Internasional Soekarno Hatta Cengkareng,"Vol.1, No.2, 2019.

Haris Kurniawan,Sarjon Defit,Sumijan, "Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal," Vol. 1, No. 2, 80 – 89, 2020.

Aline Embun Pramadhani, Tedy Setiadi, "Penerapan Data Mining Untuk Klasifikasi Prediksi Penyakit Ispa (Infeksi Saluran Pernapasan Akut) Dengan Algoritma Decision Tree (Id3),"Vol.2,No.1,160-168,2014.

Anindya Khrisna Wardhani,"Implementasi Algoritma K-Means Untuk Pengelompokkan Penyakit Pasien Pada Puskesmas Kajen Pekalongan,"Vol.14,No.1,30-37,2016.

Ahmad Mauliyadi M,Hizir Sofyan,Muhammad Subianto," Perbandingan Metode Fuzzy C-Means (Fcm) Dan Fuzzy Gustafson-Kessel (Fgk) Menggunakan Data Citra Satelit Quickbird," Jurnal Transenden 2,1–5,2013.

Badan Pusat Statistik Provinsi Sumatera Selatan,"Laporan Perekonomian Provinsi Sumatera Selatan 2023,"Vol.5,2024

Wahyudi Kumorotomo," Akuntabilitas Anggaran Publik: Isu Politik, Prioritas Belanja Dan Silpa Dalam Alokasi Apbd Di Beberapa Daerah," Konferensi Administrasi Negara Ke-3,2010.

Atakishiyev, S., Reformat, M. Z." Analysis Of Word Embeddings Using Fuzzy Clustering," Springer International Publishing, Pp.539-551,2021.

Badan Pusat Statistik Provinsi Sumatera Selatan,"Statistik Daerah Provinsi Sumatera Selatan ,"2023.

Dey,Subhrasankha,Tanmoy Dam. "Rainfall-Runoff Prediction Using A Gustafson-Kessel Clustering Based Takagi-Sugeno Fuzzy Model," 2021 Ieee Symposium Series On Computational Intelligence (Ssci). Ieee, 2021.

Hosein Hashemi, Abdolrahim Javaherian And Robert Babusk," A Semi-Supervised Method To Detect Seismic Random Noise With Fuzzy Gk Clustering," Iop Publishing,2008