JISTech (Journal of Islamic Science and Technology)
JISTech, 9(1), 53-57, Januari – Juni 2024
ISSN: 2528-5718
http://jurnal.uinsu.ac.id/index.php/jistech

FUZZY C-MEANS CLUSTERING TECHNIQUE ANALYSIS OF NORTH SUMATRA PROVINCE'S DISTRICT/CITY CLASSIFICATION BASED ON COMMUNITY SOCIAL WELFARE LEVEL

Rizky Indri Syahputri
Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
Email : rizkyindrisyahputri@gmail.com

Abstrak

Penelitian ini bertujuan untuk menggambarkan signifikansi dan alasan di balik pemilihan -n kelompok terbaik untuk mengatur kabupaten dan kota di Provinsi Sumatera Utara menurut Indeks Kesejahteraan Rakyat 2022 menggunakan Fuzzy C-Means. Penelitian ini menggunakan data sekunder dari situs web BPS Provinsi Sumatera Utara, dengan menggunakan variabel-variabel seperti kepadatan penduduk, rasio ketergantungan, rata-rata lama sekolah, pengeluaran rata-rata dalam juta rupiah, dan tingkat pengangguran. Data dari penelitian ini akan berfungsi sebagai landasan untuk keputusan mengenai kebijakan dan solusi untuk masalah yang dihadapi setiap kelompok. Penelitian ini menggunakan fuzzy c-means untuk mengelompokkan data di Provinsi Sumatera Utara. Algoritma tersebut melibatkan penentuan jumlah cluster, daya, iterasi maksimum, kesalahan harapan terkecil, dan fungsi tujuan awal. Angka acak dibangkitkan dan matriks partisi dihitung. Hasilnya menunjukkan dua cluster, satu untuk kota dengan tingkat pengangguran dan kepadatan penduduk yang tinggi, dan satu lagi untuk daerah dengan sekolah dan pengeluaran yang rendah. Pemerintah harus memprioritaskan penanganan kepadatan penduduk dan pengangguran yang tinggi di Klaster 1, sementara menangani rasio ketergantungan dan lama sekolah di Klaster 2. Temuan ini dapat menjadi panduan dalam pengambilan keputusan program kerja di masa mendatang.

Kata Kunci*:* C-Means, Klaster, Fuzzy, Tingkat Kesejahteraan Sosial Masyarakat.

Abstract

This study uses Fuzzy C-Means to group districts and cities in North Sumatra Province based on the 2022 People's Welfare Index. It tries to explain the significance of this process and the rationale behind the cluster selection. Utilizing factors including population density, dependence ratio, average length of schooling, average spending in million rupiah, and poverty level, this study employs secondary data from the BPS website of North Sumatra Province. The information obtained from this study will be used to inform decisions about policies and how to address issues that each group faces. In this work, data from the North Sumatra Province is clustered using fuzzy c-means. The algorithm involves figuring out the initial objective function, power, maximum iteration, least predicted error, and number of clusters. The partition matrix is computed and random numbers show up. Two clusters are identified by the results: one for cities with high rates of poverty and population density, and another for regions with low levels of education and spending. While addressing the dependence ratio and length of schooling in Cluster 2, the government should give priority to the high population density and unemployment in Cluster 1. Future work program decisions can be made using these findings as a guide.

Keywords: *C-Means, Claster, Fuzzy, Community Social Welfare Level.*

Introduction

The clustering method (sometimes called "clustering") divides data into many groups or clusters according to predefined similarities, with a high degree of resemblance or features and a notable degree of diversity between each group or cluster [1]. Data with these variations and similarities can be used to calculate distance. Among the clustering techniques is the Fuzzy C-Means method. North Sumatra Province's 1945 Constitution states in Paragraph 4 that the state's mission is to prosper its citizens [2]. The necessities of human life evolve with the passage of time and with the times. Being materially and spiritually sufficient is a state that might be termed prosperous. A community is deemed affluent when its members are able to achieve their basic necessities [3]. Welfare can be measured using a number of indices, such as district/city GRDP, population density, number of impoverished, purchasing power of the populace, number of workers, life expectancy, literacy rate, average and expected length of schooling, and open unemployment rate [4]. There are 33 cities and districts in North Sumatra. The welfare state differs from city to district. In the North Sumatra Province, districts and cities can be grouped together to promote equity and enhance the well-being of the populace. This classification is intended to enable the implementation of government objectives and programs in accordance with the limitations in each cluster. To combine them, this needs to be based on the people's welfare index [5]. Fuzzy C-Means is one of the clustering techniques (FCM). FCM, which was first presented by Jim Bezdek in 1981, assesses a cluster's degree of membership by considering whether data is present at every location. In the beginning, when the cluster center is imprecise, this is done. Clustering difficulties in astronomy, target recognition, picture segmentation, chemistry, geology, and medical diagnostics can be resolved with this approach [6],[7]. This method's objective is to gather information and identify the cluster center [8].

The foundation of this study will be the Fuzzy C-Means (FCM) algorithm. Each piece of data in a cluster in the FCM clustering technique is determined by its membership value. Find the group's center before inserting the point into the group. The cluster center membership value must be determined frequently to guarantee that the cluster center will relocate to the correct position because the circumstances and points of each data connected to the degree of membership of the cluster center may not be particularly accurate [9]. The goal of this case study is to illustrate the significance and rationale behind choosing the best n groups to organize districts and cities in the province of North Sumatra according to the 2022 People's Welfare Index, using Fuzzy C-Means. It is believed that the data from this study will be beneficial and serve as a foundation for choices concerning policies and solutions that are suitable for the problems facing each group.

Research Methods

1. Research Data

Secondary data from the North Sumatra Province's BPS website was used for this article's study. Using variables chosen by the researcher, the approach used is clustering using the Fuzzy C-Means algorithm with the goal of organizing Central Java Regency/City into multiple clusters. Population density (X1), dependency ratio (X2), average length of schooling (years) (X3), average expenditure in million rupiah (RPD) (X4), and unemployment rate (X5) are the variables that are used. Table 1 contains the research variables.

Table 1. Research Variable

Districts/Cities	X_1	X_2	X_3	X_4	X_5
Asahan	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,4}$	$x_{1,5}$
Batu Bara	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	$x_{2,4}$	$x_{2,5}$
Binjai	$x_{3,1}$	$x_{3,2}$	$x_{3,3}$	$x_{3,4}$	$x_{3,5}$
…	…	…	…	…	…
Tebing Tinggi	$x_{32,1}$	$x_{32,2}$	$x_{32,3}$	$x_{32,4}$	$x_{32,5}$
Toba	$x_{33,1}$	$x_{33,2}$	$x_{33,3}$	$x_{33,4}$	$x_{33,5}$

2. Data Processing

The steps taken in data analysis are as follows:

1) Grouping data using the fuzzy c-means method. The following are the stages in the fuzzy c-means algorithm:

a. Determining:
   i. The number of clusters (c) is the number of clusters to be formed.
   ii. The power (m) is generally $1.25 \leq m \leq 2$, where the greater the value of m used, the more the data center is in the middle of the data.
   iii. Maximum iteration (MaxIter) is a loop that will stop if the maximum iteration value has been reached.
   iv. The smallest expected error (ε) is the value limit that makes the loop end after the expected error value is obtained.
   v. The initial objective function ($P0$) is a function that will be optimized (maximum or minimum),
      1) value of 0 means to get the minimum value.
      2) Generating random numbers μik, i = 1, 2, …, n; k = 1, 2, …, c as elements of the initial partition matrix U of size $i \times k$.
      3) Calculating the kth cluster center: $v$kj, where k = 1, 2, …, c and j = 1, 2, …, p using the following equation:

$$v_{kj} = \frac{\sum_{i=1}^{n}((\mu_{ik})^{m} x_{ij})}{\sum_{i=1}^{n}(\mu_{ik})^{m}}$$

Dimana

μ_{ik} : Membership value on the i-th sample data and k-th cluster

x_{ij} : The i-th sample data and the j-th variable

m : Rank

b. Calculate the objective function at the t-th iteration, Pt using the following equation:

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{p}\left(x_{ij} - v_{kj}\right)^2\right](\mu_{ik})^m\right)$$

where,

x_{ij} = The i-th sample data and the j-th variable
V_{kj} = Midpoint between the k-th variable and the j-th variable
μ_{ij} = Membership values for the i-th sample data and k-th cluster.
m = Rank

c. Calculate the changes in the membership matrix using the following equation:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{p}\left(x_{ij} - v_{kj}\right)^2\right]^{\frac{-1}{m-1}}}{\sum_{j=1}^{c}\left[\sum_{j=1}^{p}\left(x_{ij} - v_{kj}\right)^2\right]^{\frac{-1}{m-1}}}$$

di mana,

x_{ij} = The i-th sample data and the j-th variable
V_{kj} = Midpoint between the k-th variable and the j-th variable
m = Rank

d. Repeat stepsa a to c until ($|P$t$-P$t-1$|<\boldsymbol{\varepsilon}$) or t>maxiter is met.
e. Repeat steps a to d for different clusters.
f. Calculate the validity index of the Partition Coefficient (PC) for each cluster using the following equation:

$$PC(c) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{c}(\mu_{ik})^2$$

Where, µik is the final membership value in the i-th sample and k-th cluster.

2) Interpreting optimal cluster results

Result and Discussion

1. The Calculation Results

Descriptive Analysis in this study, iterations were carried out 4 times in its completion. The results of the iteration are in Table 2.

Tabel 2. Descriptive Analysis

Districts/Cities	X_1	X_2	X_3	X_4	X_5
Asahan	2385.141336	7.578463123	0.044084015	1.776490402	4.379889938
Batu Bara	48934.78929	1.814676104	0.291558892	0.221700174	3.432936982
Binjai	8824833.26	10.13085786	4.622663677	1.220694583	4.296577192
…	…	…	…	…	…
Tebing Tinggi	18525122.65	11.71625027	3.312538554	8.543054928	4.896566416
Toba	28977.54093	11.74500854	2.402618	4.901133947	7.437518069

2. Clusters are created as a result.

With the implementation of FMC in districts/cities in the province of North Sumatra, the results of the clusters formed are shown in the table.

Tabel 3. Cluster as result

Cluster	Districts/Cities	Characteristics
I	Binjai, Medan, Pematangsiantar, Sibolga, Tanjungbalai, Tebing Tinggi	High unemployment rate and great population density
II	Asahan, Batu Bara, Dairi, Deli Serdang, Gunungsitoli, Humbang Hasundutan, Karo, Labuanbatu Utara, Labuhan Batu, Labuhanbatu Selatan, Langkat, Mandailing Natal, Nias, Nias Barat, Nias Selatan, Nias Utara, Padang Lawas, Padang Lawas Utara, Padangsidimpuan, Pakpak Bharat, Samosir, Serdang Bedagai, Simalungun, Tapanuli Selatan, Tapanuli Tengah, Tapanuli Utara, Toba.	The dependency ratio is high, although the average years of schooling and average expenditure are both low.

Conclusion

According to the validity results obtained from the FCM technique, it has been determined that the most suitable number of clusters for grouping districts and cities in the province of North Sumatra, based on the community welfare index, is 2 clusters. Cluster 1 comprises the cities of Binjai, Medan, Pematangsiantar, Sibolga, Tanjungbalai, and Tebing Tinggi. The members of Cluster 2 include Asahan, Batu Bara, Dairi, Deli Serdang, Gunungsitoli, Humbang Hasundutan, Karo, North Labuanbatu, Labuhan Batu, South Labuhanbatu, Langkat, Mandailing Natal, Nias, West Nias, South Nias, North Nias, Padang Lawas, North Padang Lawas, Padangsidimpuan, Pakpak Bharat, Samosir, Serdang Bedagai, Simalungun, South Tapanuli, Central Tapanuli, North Tapanuli, and Toba.

Regarding Cluster 1, the government should prioritize addressing issues related to high population density and unemployment. Cluster 2 highlights the need for the government to address the significant dependency ratio and the below-average length of schooling and expenditure. The variables and locations within cluster 1 and cluster 2 can serve as a point of reference for the North Sumatra Provincial Government in making decisions regarding the focus of their work program, with the aim of improving the region in the future.

References

[1] Kastawan Putu Wirya (2018). Implementasi Algoritma C5.0 pada Penilaian Kinerja Pegawai Negeri Sipil. Jurnal Majalah Ilmiah Teknologi Elektro Vol. XVII

[2] Kholis, N., Kunci, K., Kesejahteraan, :, Islam, J., & Sosial, D. E. (n.d.). Kesejahteraan SosialDi Indonesia Perspektif Ekonomi Islam.

[3] Setiawan, H. H. (2019). Merumuskan Indeks Kesejahteraan Sosial (Iks) Di Indonesia. Sosio Informa, 5(3). https://doi.org/10.33007/inf.v5i3.1786.

[4] Alwi, W., & Hasrul, M. (2018). Analisis Klaster Untuk Pengelompokkan Kabupaten/KotaDi Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat. Jurnal MSA(Matematika Dan Statistika Serta Aplikasinya), 6(1), 35.https://doi.org/10.24252/msa.v6i1.4782

[5] Yulianto, S., & Hidayatullah, K. H. (2016). Analisis Klaster Untuk Pengelompokan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Rakyat. Statistika, 2(1), 56–63. https://jurnal.unimus.ac.id/index.php/statistik/article/view/1115

[6] Riandayani, D. A., Darma Putra, I. K. G., & Buana, P. W. (2014). Comparing Fuzzy Logicand Fuzzy C-Means (FCM) on summarizing indonesian language document. Journal of Theoretical and Applied Information Technology, 59(3), 718–724.

[7] Suroso, D. J., Cherntanomwong, P., Sooraksa, P., & Takada, J. I. (2011). Location fingerprint technique using Fuzzy C-Means clustering algorithm for indoor localization.IEEE Region 10 Annual International Conference, Proceedings/TENCON, November, 88–92. https://doi.org/10.1109/TENCON.2011.6129069

[8] Simbolon, C. L., Kusumastuti, N., & Irawan, B. (2013). Clustering lulusan mahasiswa matematika fmipa untan pontianak menggunakan algoritma fuzzy c - means. Buletin Ilmiah Mat. Stat. Dan Terapannya (Bimaster), 02(1), 21–26.

[9] Mulyani, S., Sari, B. N., & Ridha, A. A. (2020). Clustering Productivity of Rice in Karawang Regency Using the Fuzzy C-Means Method. Indonesian Journal of Artificial Intelligence andData Mining, 3(2), 103–112. https://ejournal.uin-suska.ac.id/index.php/IJAIDM/article/view/10415