JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 153 – 160 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3602
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

APPLICATION OF THE K-MEANS METHOD FOR GROUPING COMMUNITY WELFARE LEVELS IN CENTRAL JAVA PROVINCE

Taufik Hidayat[1*], Yuni Handayani[1], Dian Novitaningrum[1]
[1]Teknik Informatika, Universitas Selamat Sri
email: *taufikhidayat.jc@gmail.com

Abstract: Welfare is one of the things that determines the progress of a region, to achieve the welfare of its people, especially in the economic sector, a technique is needed to measure welfare that continues to change. This study aims to analyze the differences in the level of community welfare in Central Java Province by grouping regions based on several indicators. Grouping is done using data from various sources that include the main indicators of welfare. The method used in this study uses the K-Means data mining algorithm to group regional data according to their level of welfare. The results of the analysis divide the regions into three categories: Medium Welfare Level, which includes Banyumas, Purworejo, Boyolali, Klaten, Sukoharjo, Karanganyar, Sragen, Kudus, Jepara, Demak, Semarang, Kendal, and Pekalongan City and Tegal City, High Welfare Level, consisting of Magelang City, Surakarta City, Salatiga City, and Semarang City; and Low Welfare Level, covering Cilacap, Purbalingga, Banjarnegara, Kebumen, Wonosobo, Magelang, Wonogiri, Grobogan, Blora, Rembang, Pati, Temanggung, Batang, Pekalongan, Pemalang, Tegal, and Brebes Regencies. The findings show that the C2 region has a longer average length of schooling, higher per capita expenditure, and better HDI, reflecting a higher quality of life. This study provides an overview of welfare inequality in Central Java Province and suggests the need for more focused policies to improve the quality of life in each category of region.

Keywords: clustering; k-means; welfare

Abstrak: Kesejahreraan merupakan salah satu hal yang menentukan kemajuan suatu wilayah, untuk mencapai kesejahteraan masyarakatnya terutama di bidang ekonomi di perlukan teknik untuk mengukur kesejahteraan yang terus berubah, Penelitian ini bertujuan untuk menganalisis perbedaan tingkat kesejahteraan masyarakat di Provinsi Jawa Tengah dengan mengelompokkan wilayah berdasarkan beberapa indikator. Pengelompokan dilakukan menggunakan data dari berbagai sumber yang mencakup indikator-indikator utama kesejahteraan. Metode yang di gunakan dalam penelitian ini menggunakan algoritma data mining K-Means untuk mengelompokkan data wilayah menurut tingkat kesejahteraannya. Hasil analisis membagi wilayah menjadi tiga kategori: Tingkat Kesejahteraan Sedang, yang mencakup Kabupaten Banyumas, Purworejo, Boyolali, Klaten, Sukoharjo, Karanganyar, Sragen, Kudus, Jepara, Demak, Semarang, Kendal, serta Kota Pekalongan dan Kota Tegal, Tingkat Kesejahteraan Tinggi, terdiri dari Kota Magelang, Kota Surakarta, Kota Salatiga, dan Kota Semarang; dan Tingkat Kesejahteraan Rendah, mencakup Kabupaten Cilacap, Purbalingga, Banjarnegara, Kebumen, Wonosobo, Magelang, Wonogiri, Grobogan, Blora, Rembang, Pati, Temanggung, Temuan menunjukkan bahwa wilayah C2 memiliki rata-rata lama sekolah yang lebih panjang, pengeluaran per kapita yang lebih tinggi, dan IPM yang lebih baik, mencerminkan kualitas hidup yang lebih tinggi. Penelitian ini memberikan gambaran tentang ketidakmerataan kesejahteraan di Provinsi Jawa Tengah dan menyarankan perlunya kebijakan yang lebih terfokus untuk meningkatkan kualitas hidup di setiap kategori wilayah.

Kata Kunci: clustering; k-means; kesejahteraan

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 153 – 160 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3602
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

Each stage of economic development aims to achieve public welfare. In Indonesia, the fourth paragraph of the opening of the 1945 Constitution states that the main goal of the state is to achieve public welfare. To achieve this goal, the government has carried out various development programs to achieve public welfare. However, economists are still investigating how to convert welfare quantitatively [1]. Over the past five decades, the understanding and techniques for measuring economic welfare have continued to change. Welfare is the mainstay of the hopes and noble ideals of the struggle of the Indonesian people. Welfare is also something that determines the progress of a region.

A person's ability to meet their basic needs, such as clothing, food, housing, school, and medical care, is called welfare. Family welfare is the best way to measure a person's welfare [2]. Family welfare is when all family members live in harmony and fulfill their physical and social needs, without experiencing serious obstacles, and family members can work together to solve problems together, resulting in a standard of living for the family [3].

The level of community welfare is an important indicator that reflects the quality of life and economic conditions of community groups. However, there are various problems that affect the assessment and efforts to improve this welfare. One of the main issues is the uneven distribution of wealth, which creates significant gaps between rich and poor groups and between more developed and less developed regions. In addition, limited data, both in terms of accuracy and timeliness, often hinders a proper understanding of community conditions [4].

Poverty and limited access to basic services such as health and education are also major problems, affecting many individuals who live in poor conditions. Poor health and nutrition, with the prevalence of disease and malnutrition, have a direct impact on welfare. Low education and inadequate quality of education limit employment opportunities and income, and reduce the benefits of education [5].

Inadequate housing and poor infrastructure also worsen the quality of life. In addition, social and psychological issues such as mental health, violence, and environmental insecurity also play a significant role in people's well-being. Ineffective policies, corruption, and limited resources often hamper efforts to improve well-being. Rapid social and economic changes, including globalization and environmental change, add to the complexity of the challenges faced [6].

To address these problems, improvements in data collection and analysis are needed to obtain a more accurate picture. Well-targeted social programs and transparent policy reforms as well as investment in infrastructure and basic services are essential. Focusing on improving education and health is also a crucial step to support people's well-being more effectively [7].

Therefore, it is necessary to measure the level of community welfare in order to have a standard level of welfare that is based on, and one of the objectives of this study is to measure the level of community welfare based on indicators that greatly influence the level of welfare from the many factors that influence it, and also refer to several previous studies that support it.

There are several studies related to the level of community welfare, including a journal entitled The Level of Community Welfare of Fishermen in Benua Baru Ilir Village Based on the Central Statistics Agency Indicators, where in this journal the results were obtained that 15% of respondents were included in the category of families with a high level of welfare, while the other 85% were in the category of families with a moderate level of welfare, according to the indicators set by the Central Statistics Agency [1] and there is also a study entitled Application of the K-Means Algorithm in Grouping Community Welfare in Karawang Regency, it was concluded that Cluster 1 is a group with a high level of community welfare and covers 7 sub-districts. Cluster 2 reflects a group with a moderate level of community welfare, consisting of 8 sub-districts. Meanwhile, Cluster 3 shows a group with a low level of community welfare and covers 7 sub-districts. The clustering process using 3 clusters produces an SSE value of 23.58788 with an accuracy level reaching 72.9% [4].

To date, the Indonesian government has continued to focus on the distribution of people's well-being. The concept of people's well-being covers various aspects of life, so it can be used as one measure of a country's progress. Therefore, cluster analysis is needed to group districts in Central Java Province based on community welfare metrics. Thus, this will help the government make policies related to people's welfare easier to implement [8].

Grouping regions according to community welfare levels aims to understand and manage differences in living conditions in various areas more effectively. By grouping regions based on welfare levels, we can identify the specific needs of each region, so that social programs and policies can be tailored to those specific needs [9].

This also allows for more efficient planning and allocation of resources, ensuring that budgets and assistance are focused on areas that need it most. In addition, regional grouping facilitates the evaluation of policies and programs by comparing impacts at various levels of welfare, and helps map existing social and economic inequalities. This is important for developing regional development strategies that are in accordance with local characteristics, improving the quality of public services, and formulating evidence-based policies [10].

Thus, regional grouping can also raise awareness of inequality and encourage community participation in the planning and implementation of development programs, supporting efforts to reduce inequality and improve welfare.

METHOD

The approach used in this study is quantitative, focusing on the collection and analysis of statistical data related to indicators of community welfare, such as per capita expenditure, Human Development Index, and average length of schooling. This approach aims to identify patterns and characteristics of welfare in various regions of Central Java Province, , the technique used in this study uses one of the data grouping algorithms in data mining, namely the K Means algorithm.

Data Mining & K-Means Method
Data Collection Techniques

Data were collected through secondary data from statistical institutions and the Kendal Regency government, namely the Central Java Provincial BPS. The data used are related to average per capita expenditure, human development

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 153 – 160 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3602
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

index, and average length of schooling in 2023, these three variables are used as the basis for applying the clustering algorithm using K-Means .

Data Preparation

Furthermore, the average calculation is used to combine the three initial data into two variables: per capita expenditure, human development index, and average length of schooling.

Normalisasi Data

Data normalization is a technique for transforming data values into a uniform scale. This is important to ensure that each variable has a comparable contribution in the analysis, especially when the data is used in a scale-sensitive algorithm, such as clustering or regression analysis (6).
The following formula will be used in normalizing the available indicator data:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Description:
x: Original data value.
x': Data value after normalization.
x min: Minimum value in the data.
x max: Maximum value in the data.

K-Means

K-Means is one of the non-hierarchical or Partitional Clustering methods. The K-Means algorithm attempts to group existing data into several clusters, where the data within a cluster have similar characteristics with one another and have different characteristics from the data in other clusters. Below is the flowchart of the K-Means Clustering algorithm [11].

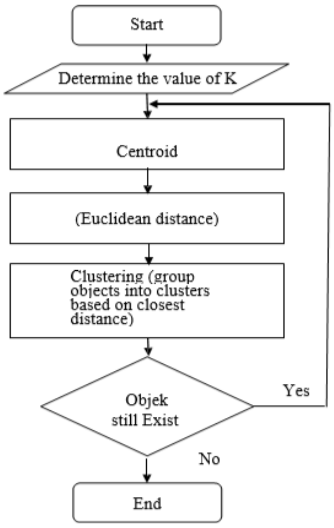


Image 1. Flowchart of the K-Means Clustering Algorithm

At the K Means Algorithm stage there are several stages and at the cluster distance testing stage it is necessary to calculate the distance between each data using the following equation:

$$d_{euclidean} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (2)$$

Description:
deuclidean: Euclidean distance between two points
xi: The value i-th attribute of point x.
yi: The value i-th attribute of point y.
n: Number of dimensions/attributes.

Then the calculation is done using the eclidean distance formula and if the cluster group results have been obtained from each data grouped based on their groups and the new centroid value is searched using the average formula. If the new centroid value has been obtained, the calculation is re-done using the eclidean distance formula and the data from each data group must be the same

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 153 – 160 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3602
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

as the previous group if there is a change in the group then the stage must be repeated again, until the group results do not change.

RESULTS AND DISCUSSION

Determination of the Number of Clusters.

The K-Means algorithm groups data into several groups (clusters) based on similarities between groups. This is done by calculating the centroid, or centroid, of each group, and assigning each data point to the group with the closest centroid.

Determining the ideal number of clusters for a method like K-Means is an important step. The number of clusters or groups is calculated after the numbers for each variable are normalized. This research will divide them into three groups, with the central value of each group in the first iteration selected randomly.Cluster distance testing.

Although there are many methods that can be used to identify initial centroids as a first step, the value centroids of each cluster can also be calculated randomly. Calculate the distance between each data and the nearest cluster. The distance between two objects determines their proximity to each other. In the same way, the distance between the data and the cluster center, or centroid, is determined,

Apart from that, the test was carried out 12 times due to the large amount of data that had to be grouped. However, the use of K-Means in R studio relies heavily on centroid identification, so randomization occurs frequently and it is difficult to obtain unique initial results. Thus, randomization (nstart) was carried out 12 times.

Table 1. Cluster Results

data ke-i	Distance c1	Distance c2	Distance c3	cluster
1	0,105659	1,17221	0,37287	1
2	0,311832	0,969738	0,186017	3
3	0,046761	1,265815	0,467235	1
4	0,175172	1,428281	0,628957	1
5	0,186556	1,259274	0,491824	1
6	0,29981	0,987623	0,229948	3
7	0,158758	1,296046	0,505212	1
8	0,107447	1,207656	0,421042	1
9	0,508522	0,82084	0,190577	3
10	0,629968	0,626182	0,182431	3
11	0,712149	0,594715	0,310244	3
12	0,107465	1,229005	0,442621	1
13	0,54251	0,726414	0,136463	3
14	0,454875	0,878233	0,196333	3
…	………..	…………	…………	..
35	0,694706	0,57823	0,249452	3

The following are the results of clustering using the K-Means method:

Table 2. Cluster Division

Cluster	Number Of Members
c1	17
c2	4
c3	14

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 153 – 160 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3602
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Table 3 Final Cluster Results

Cluster	Regency/City
Cluster 1	Kab. Cilacap, Kab. Purbal-ingga, Kab. Banjarnegara, Kab. Kebumen, Kab. Wono-sobo, Kab. Magelang, Kab. Wonogiri, Kab. Grobogan, Kab. Blora, Kab. Rembang, Kab. Pati, Kab. Te-manggung, Kab. Batang, Kab. Pekalongan, Kab. Pem-alang, Kab. Tegal, Kab. Brebes
Cluster 2	Kota Magelang, Kota Sura-karta, Kota Salatiga, Kota Semarang
Cluster 3	Kab. Banyumas, Kab. Pur-worejo, Kab. Boyolali, Kab. Klaten, Kab. Sukoharjo, Kab. Karanganyar, Kab. Sragen, Kab. Kudus, Kab. Jepara, Kab. Demak, Kab. Sema-rang, Kab. Kendal, Kota Pekalongan,Kota Tegal

Final centroid profiling, which is an average calculation of pre-normalized data, is required to obtain an interpreta-tion of the above grouping vector.

the calculation process continues to be carried out repeatedly if the cen-troid value changes, then a new centroid value is searched by calculating the mid-point and calculated again using the Eu-clidean distance formula, and the data is grouped again into the same cluster or not, and after the calculation is carried out long enough to get the final centroid and cluster values that do not change

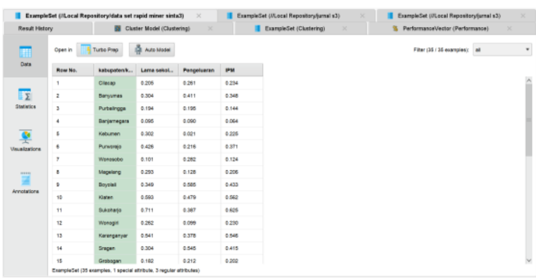
Tabel 4. Final Centroid Value

Cluster	Average Length of Schooling	Per capita expendi-ture	HDI
Cluster 1	7,39	10712,29	70,9876
Cluster 2	11,0625	15528,75	83,5325
Cluster 3	8,615714	12616,5	75,6971

The following is an explanation of the final centroid value:

1. C3 is a group category with a me-dium level of community welfare: This group consists of Kab. Banyumas, Kab. Purworejo, Kab. Boyolali, Kab. Klaten, Kab. Su-koharjo, Kab. Ka-ranganyar, Kab. Sragen, Kab. Kudus, Kab. Jepara, Kab. Demak, Kab. Se-marang, Kab. Kendal, Pekalon City, Tegal City
2. C2 is a group category with a high level of community welfare: This group consists of Magelang City, Surakarta City, Salatiga City, Semarang City.
3. C1 is a group category with a low level of community welfare: This group consists of Kab. Cilacap, Kab. Purbal-ingga, Kab. Ban-jarnegara, Kab. Kebumen, Kab. Wonosobo, Kab. Magelang, Kab. Wonogiri, Kab. Gro-Bogan, Kab. Blora, Kab. Rembang, Kab. Pati, Kab. Temanggung, Kab. Batang, Kab. Pekalongan, Kab. Pema-lang, Kab. Tegal, Kab. Brebes

The following is a display of the calculation results using the system, the results obtained are the same as the pre-vious calculation.



JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 153 – 160 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3602
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Image 2. Calculation results with the system

On an analysis of regional groupings in Central Java Province according to the level of community welfare, it can be seen that there are significant differences in the average length of schooling, per capita expenditure, and the Human Development Index (HDI) in various regional groups.

CONCLUSION

Regions with a high level of prosperity (C2), namely Magelang City, Surakarta City, Salatiga City, and Semarang City, generally have a longer average length of schooling, higher per capita expenditure, and better HDI. . This reflects better access to education, stronger economic capabilities, and a higher quality of life compared to other regions.

On the other hand, areas with a moderate level of prosperity (C3), such as Banyumas, Purworejo, Boyolali, Klaten, and surrounding districts, show relatively lower average years of schooling and per capita expenditure compared to the C2 group, although they are still better than groups with low welfare. The HDI in this group is also at the middle level, indicating that despite progress, challenges in educational and economic access still exist.

In the group with a low level of prosperity (C1), which includes the regencies of Cilacap, Purbalingga, Banjarnega, and others, the average length of schooling tends to be shorter, per capita expenditure is lower, and the HDI shows poor conditions. most challenging. These regions face greater difficulties in terms of access to education, income and quality of life, which impact the overall well-being of the community. In conclusion, the differences in average years of schooling, per capita expenditure, and HDI between these regional groups reflect the inequality in welfare in Central Java Province. A deeper understanding of these variables can help in designing more targeted interventions to improve well-being in each regional group.

BIBLIOGRAPHY

[1] E. Sugiharto, J. Sosial, E. P. Fpik, and U. Samarinda, "TINGKAT KESEJAHTERAAN MASYARAKAT NELAYAN DESA BENUA BARU ILIR BERDASARKAN INDIKATOR BADAN PUSAT STATISTIK (The Welfare Level of Fisherman Society of Benua Baru Ilir Village Based on Badan Pusat Statistik Indicator)," *Epp*, vol. 4, no. 2, pp. 32–36, 2007.

[2] N. A. Nabilah, H. Perdana, and E. Sulistianingsih, "Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Kesejahteraan Masyarakat Dengan Algoritma K-Means++," *Bul. Ilm. Math. Stat. dan Ter.*, vol. 13, no. 3, pp. 419–426, 2024.

[3] T. Hidayat, "Klasifikasi Data Jamaah Umroh Menggunakan Metode K-Means Clustering," *J. Sistim Inf. dan Teknol.*, vol. 4, pp. 19–24, 2022, doi: 10.37034/jsisfotek.v4i1.115.

[4] D. Fitriani, T. N. Padilah, and B. N. Sari, "Penerapan Algoritma K-Means Dalam Pengelompokan Kesejahteraan Rakyat Berdasarkan Kecamatan di Kabupaten Karawang," *Progresif J. Ilm. Komput.*, vol. 17, no. 2, p. 73,

2021, doi: 10.35889/progresif.v17i2.649.

[5] N. Nugroho and F. D. Adhinata, "Penggunaan Metode K-Means dan K-Means++ Sebagai Clustering Data Covid-19 di Pulau Jawa," *Teknika*, vol. 11, no. 3, pp. 170–179, 2022, doi: 10.34148/teknika.v11i3.502.

[6] S. Wulandari and D. Novita, "Analisis Clustering Virus MERS-CoV Menggunakan Metode Spectral Clustering Dan Algoritma K-Means," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 5, no. 3, p. 315, 2021, doi: 10.30998/string.v5i3.7942.

[7] T. Nanda Khofifah and R. Fajriyah, "Perbandingan Dampak Bencana Angin Kencang Tahun 2020 Dan 2021 Daerah Istimewa Yogyakarta Berdasarkan Metode K-means Clustering," *Emerg. Stat. Data Sci. J.*, vol. 2, no. 1, pp. 107–121, 2024, doi: 10.20885/esds.vol2.iss.1.art11.

[8] S. Hanifah and A. H. Primandari, "Implementasi Metode K-Means Clustering dalam Pengelompokan Kabupaten/ Kota di Provinsi NTB Berdasarkan Indikator Pendidikan," *Emerg. Stat. Data Sci. J.*, vol. 1, no. 3, pp. 378–393, 2023, doi: 10.20885/esds.vol1.iss.3.art44.

[9] S. H. Widiastuti and R. Jumardi, "Pengelompokan Daerah Rawan Demam Berdarah dengan Metode K-Means Clustering," *J. Inf. dan Teknol.*, vol. 4, no. 4, pp. 185–190, 2022, doi: 10.37034/jidt.v4i4.213.

[10] D. Ramdhan, G. Dwilestari, R. D. Dana, A. Ajiz, and K. Kaslani, "Clustering Data Persediaan Barang Dengan Menggunakan Metode K-Means," *MEANS (Media Inf. Anal. dan Sist.*, vol. 7, no. 1, pp. 1–9, 2022, doi: 10.54367/means.v7i1.1826.

[11] M. Yunus, "Metode Clustering K-Means Pada Penjualan Handphone," vol. 4, no. 2, pp. 75–87, 2024.