CLUSTERIZATION OF DISTRICTS AND CITIES IN JAMBI PROVINCE BASED ON PUBLIC HEALTH INDICATORS USING THE K-MEANS METHOD

Nayla Desviona[1*], Marwah Masruroh[2], Rifki Chandra Utama[3]

[1]Departement of Management, Universitas Muhammadiyah Jambi, Province of Jambi, Indonesia
[2]Departement of D3 Mechanical Engineering, Politeknik Negeri Jakarta, Province of Jakarta, Indonesia
[3]Departement of Actuaria, Institut Teknologi dan Bisnis Muhammadiyah Purbalingga, Province of Jawa Tengah, Indonesia
*Correspondence: nayladesviona@umjambi.ac.id

ABSTRACT
Health is an important foundation for Indonesia, and public health is critical to improving the quality of human resources, overcoming poverty, and supporting development. In 2022, the morbidity rate in Jambi Province reached 12.11%, higher than the previous year which was only 7.16%. The morbidity rate of the people of Jambi Province in 2022 increased by around 5% compared to last year. Based on the data on the morbidity rate of the people of Jambi Province above, it can be seen that the health condition of the people of Jambi Province has decreased compared to the previous year. By using the k-means method, this study aims to group the districts and cities of Jambi Province based on their health indicators as an effort to prevent a decline in health status in the coming years. As an effort to set priorities in improving public health in Jambi Province. The research resulted in three health clusters based on health indicator data. Cluster 1, which contains Sungai Penuh City, is at a high level, meaning that health conditions in the city are very good compared to other cities. Cluster 2, which consists of West Tanjung Jabung Regency and Jambi City, is at a medium level, meaning that health conditions in this city are quite good compared to other cities. Cluster 3 Kerinci, Merangin, Sarolangun, Batanghari, East Tanjung Jabung, Tebo, and Bungo are at the lowest level, meaning that health conditions in this cluster are poor and require more attention.
Keywords: Cluster, Health, K-means Method

PRELIMINARY

Health is one of the important foundations of a country and plays a crucial role in people's role in people's lives (Subargus, 2022). Health includes physical, mental, and environmental well-being that enables everyone to live productively both socially and economically (Choi et al., 2019). Good individual health conditions have a real positive impact on a person's ability to work and increase productivity, thereby contributing to the welfare of their families and communities. In addition, a physically and mentally healthy

society plays an important role in encouraging development and economic growth in Indonesia as an effort to reduce poverty.

The public health conditions in Indonesia are influenced by various interrelated and complex factors, including healthy or unhealthy environmental conditions, individual behavior in maintaining their health, and the quality of available health services. All of these factors have a significant impact on the overall quality of life of the community (Mubarak & Kholijah, 2023).

In Indonesia, public health is very important to enhance the caliber of human capital, overcome poverty, and support development. Good individual health will affect their abilities and productivity(Japar et al., 2024). In general, public health conditions in Indonesia are influenced by various factors, including environmental conditions, individual behavior, and the quality of health services (Jannati, 2022).

Public health and welfare is a key focus of the government's initiatives aimed at enhancing national health, with the primary objective of achieving a healthy and competitively strong nation. Health is not solely a personal duty; it is also a crucial element that must be incorporated into development policies due to its significant role in fostering the growth of quality human resources (Prakoso et al., 2023). Consequently, advancement in the health sector is not merely a fundamental necessity, but also a strategic measure to enhance public services and general health to foster a healthier, more productive, and prosperous society (Hafizurrachman, 2020).

Jambi Province, similar to numerous other areas in Indonesia, confronts different obstacles in enhancing the quality of public health across all districts and cities within its jurisdiction. These difficulties pertain not just to supplying sufficient health services, but also to the necessity of comprehending and tackling the elements that influence public health in every particular area (Ramalisa et al., 2019). Consequently, initiatives aimed at enhancing health quality in the area necessitate a thorough and detailed examination of the diverse public health situations in every region. One method to enhance comprehension of the differences in health conditions across districts and cities is by applying clustering techniques (Hastari et al., 2023). This approach facilitates the clustering of areas according to specific relevant traits, thereby offering a more distinct understanding of the unique issues and requirements in every region (Chen, 2024). In this instance, the writer employed the K-Means Clustering method as the primary analytical tool.

This method is a non-hierarchical clustering technique that aims to group data into certain clusters so that data with similar traits are categorized together in the same cluster

(Wardono et al., 2019). K-Means is a clustering algorithm that works through an iterative process. The letter K refers to the number of clusters to be formed, the value of which is determined randomly. Meanwhile, Means refers to the temporary value that is the center of each cluster, also called the centroid. The space separating each piece of data and the centroid is determined by utilizing the Euclidean formula so that the closest space separating each piece of data and the centroid is obtained (Sari et al., 2020)

Based on data from the Jambi Province Central Statistics Agency in 2022, it was recorded that the morbidity rate in the province reached 12.11%, which shows an increase compared to the previous year which was only 7.16% (BPS, 2022). The increase in the morbidity rate of the community is around 5% compared to the previous year, which has a negative impact on the health conditions of the community in Jambi Province. With the increase in the morbidity rate, there is a decline in the level of public health in the region.

Based on the data on morbidity rates in Jambi Province, it can be seen that the health conditions of the people in this province have decreased compared to the previous year and are still relatively low compared to other provinces in Indonesia. To prevent a decline in health in the coming years, it is necessary to group districts and cities in an effort to determine priorities for improving public health in Jambi Province. Therefore, the researcher intends to group districts and cities in Jambi Province based on public health indicators, so this study is entitled "Clusterization of Districts and Cities in Jambi Province Based on Public Health Indicators Using The K-Means Method."

METHODS

This research uses a clustering method. This method is one effective way to understand the variation in health conditions across regions through clustering methods, which can group districts and cities based on specific characteristics. The K-Means Cluster method is a non-hierarchical clustering technique aimed at grouping data into specific clusters so that data with similar characteristics are grouped into the same cluster (Wardono et al., 2019). K-Means is a clustering algorithm that operates through an iterative process.

The letter K refers to the number of clusters that one wishes to form, which is determined randomly. Meanwhile, Means refers to the temporary values that serve as the center of each cluster, also known as the centroid. The space separating each piece of data point and the centroid is calculated using the Euclidean formula, resulting in the closest space separating each piece of data and the centroid (Sari et al., 2020). As previously explained, this research uses the K-Means Cluster method, which refers to the division of a group into

clusters. In its development, it consists of four stages: problem formulation, assumption testing, distance calculation, and clustering. The flow-cart in this research can be seen in Figure 1.
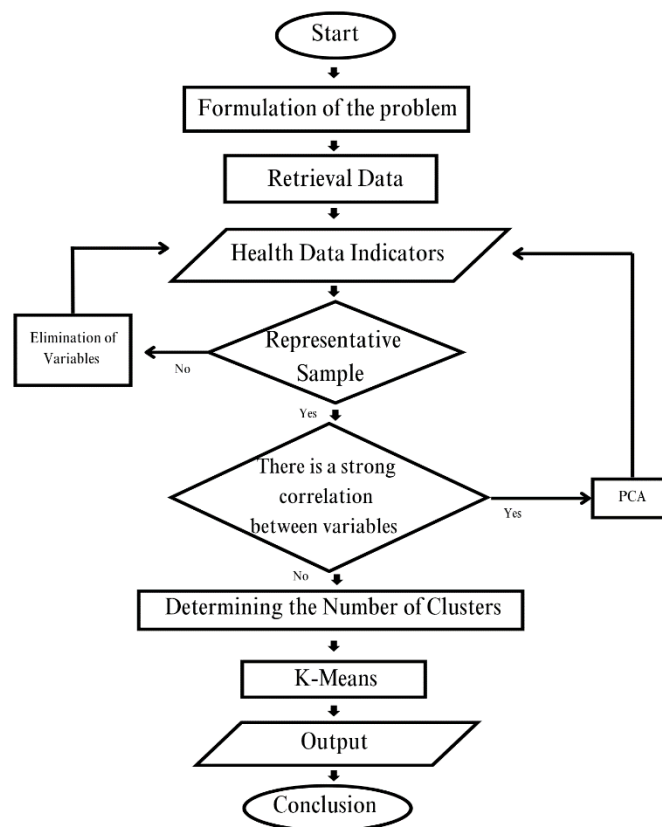

Figure 1. Flow-cart data

The subject of this research is the public health conditions in Jambi Province. The main focus of this study is to categorize the districts and cities in Jambi Province based on public health indicators. The aim is to prevent a decline in health in the coming years, as an effort to establish priorities in improving public health in Jambi Province. Here is a clarification of the development of the stages of the K-Means Cluster method, namely:

1. Formulation of The Problem

   The problem formulation processes the initial step in research or problem solving that aims to identify, define, and explain the problem to be researched or solved. This process is very important because, with a problem that is formulated clearly and precisely, researchers can determine the direction and focus of the research effectively(Ramadhani et al., 2024).

2. Assumption Testing

In cluster analysis, assumption testing is not as formal as in other inferential statistical methods (e.g., regression). However, there are some basic assumptions and steps that need to be considered to ensure that the results of the cluster analysis are valid and reliable. Here are some things to consider in the context of assumption testing in cluster analysis:

a. Representative Sample

The Kaiser-Mayer-Olkin (KMO) test is used to determine whether the sample used is sufficient for analysis. If the KMO value is in the range of 0.5 to 1, then the sample can be considered representative of the population used. According to Sauddin & Irwan, (2021), the KMO formula is as follows:

$$KMO = \frac{\sum_{i=1}^{n} \sum_{j \neq 1}^{n} r_{ij}^2}{\sum_{j \neq 1}^{n} \sum_{i \neq 1}^{n} r_{ij}^2 + \sum_{j \neq 1}^{n} \sum_{i \neq 1}^{n} (1 - r_{ij}^2)}$$

b. There is No Correlation Between Variables

It is suggested that there is no significant correlation between the independent variables used in cluster analysis. The correlation coefficient is a measuring tool that can be used to show how strong the relationship is between the variables. According to Saputra et al, (2023) the Pearson correlation coefficient value can be determined by using the formula provided:

$$r_{xiji} = \frac{\sum_{j=1}^{n} \sum_{i=1}^{n} (x_i - \bar{x}_i)(x_j - \bar{x}_j)}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x}_i)^2} \sqrt{\sum_{j=1}^{n} (x_j - \bar{x}_j)^2}}$$

Information:

$r_{x_i x_j}$ = Correlation between variables i and j

x_i, x_j = Object value on variables i and j

$\bar{x}_i, \bar{x}_j$ = Average value of variables i and j

If there is a strong correlation between independent variables, then it is necessary to handle the correlation using Principal Component Analysis (PCA).

3. Data Calculation

The distance used in this study is the Euclidean distance. According to Nishom, (2019), Euclidean distance is:

$$d_{ij} = \sqrt{\sum_{k=1}^{p} (X_{ik} - X_{ij})^2}$$

Information:

d_{ij} = The amount of space separating object i from object j

$k = 1,2,3,\ldots,p$

p = Many variables are observed

X_{ik} = The object i's value is in variable k

X_{jk} = The object j's value is in variable k

4. Clusterization

The clustering method used in this study is the k-means method. The clustering steps using the k-means method are:

a. Determining the number of clusters

b. Initialization as a centroid that can be random, according to Santoso (2014) the formula used is: k

$$d(j,k) = \sqrt{\Sigma_k^i (C_i J - C_k J)^2}$$

c. Calculating distance

d. Grouping data is derived from the nearest separation between each data point and its centroid.

$$C_{kj} = \frac{\Sigma_k^i X_{ij}}{p}$$

e. Determine the new centroid position (k)

f. Repeat processes 3 to 5 until the centroid is stable.

RESULT AND DISCUSSION

Based on the results of collecting public health data in Jambi province from the official website of the Central Bureau of Statistics in (2023) on the health statistic menu, a clusterization study using the k-means method was obtained as follows.

3.1 Assumption Testing

Step 1 Representative Sample.

To test whether the sample is suitable for analysis or not, use the Kaiser-Meyer-Olkin (KMO) test.

Table 1. Kaiser-Mayer-Olkin Test Result

Kaiser-Meyer-Olkin MSA	0.509

Based on the results of the Tabel 1 above, the KMO value obtained is 0.509, which is greater than $\alpha = 0,5$ then it can be concluded that it is rejected, which means that the representative sample or sample truly represents the population so that it is suitable for carrying out cluster analysis H_0.

Step 2 There is no Correlation Between Variables.

There should be no significant correlation between the independent variable used in cluster analysis. The correlation coefficient is a measure that can be used to indicate the strength of the relationship between variables.

Table 2. Pearson Correlation Test Result

Variables		X_1	X_2	X_3	X_4	X_5
X_1	Pearson'sr					
X_2	Pearson'sr	0.179				
X_3	Pearson'sr	-0.234	-0.455			
X_4	Pearson'sr	-0.364	-0.423	0.952		
X_5	Pearson'sr	0.476	0.100	0.389	0.351	

Based on the correlation test results table above, it can be seen that there are several variables that are correlated with each other or have a value greater than 0.05 so H_0 was not accepted, indicating. There is a correlation between variables from the data used. This can be overcome in several ways, one of which is using factor analysis with the principal component analysis method.

3.2 Principal Component Analysis

The primary objective of principal component analysis is to diminish the dimensionality of the data while retaining most of the diversity of information contained in the original variables (Beattie & Esmonde-white, 2021)
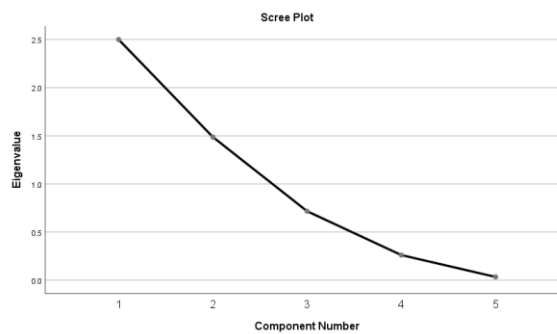

Figure 2. PCA Screen plot

The fidings of the study suggest that screen plot in Figure 2 processed using JASP software, can be seen that there are 2 main components formed because there are 2 components that have a variance value greater than 1 Therefore the health indicator data used is divided into two factors. The two factors that have been formed through the PCA results will then be used as new data that will continue to be analyzed using cluster analysis. The new data formed due to PCA is visible in the table provided:

Table 3. PCA Result Data

Distritc and City	v_1	v_2
Kerinci	1,298	0.141
Merangin	0.273	-1,722
Sarolangun	0.317	-0.684
Batanghari	-0.735	-0.128
Muaro Jambi	-0.475	-0.570
East Tanjung Jabung	0.047	-0.127
West Cape Jabung	-1,077	0.455
Tebo	0.889	-0.436
Flower	0.552	0.130
Jambi City	-2,047	0.620
Full River	0.957	2,321

After PCA is performed, assumption testing is carried out again, both KMO value testing and correlation assumption testing. First, KMO testing is carried out to see whether the sample used represents the population.

Table 4. Result of The Kaiser-Mayer-Olkin Test After PCA

*Kaiser-Meyer-Oklin*MSA	0.500

After conducting PCA, the KMO test results were obtained as 0.500, which is a figure$\geq \alpha = 0,5$. So it can be concluded that this means that the representative sample or sample has truly represented the population so that it is worthy of being analyzed cluster. Next is the correlation test as follows: H_0 rejected.

Table 5. Correlation Test Result After PCA

Variables		*Score Factor 1*	*Score Factor 2*
Score Factor 1	*Pearson's r*		
Score Factor 2	*Pearson's r*	$-3,659 \times 10^{-7}$	

Based on the correlation test result table above, it can be seen that all correlation values are less than 0.05, so H_0 is accepted, which means there is no correlation between the variables from the data used.

3.3 K-Means Method Clustering

According to the data we have collected, this is how the results of the K-Means Method Clustering will be presented.

Table 6. Result of Grouping 3 Clusterest

Cluster	Cluster Members	Number of Members
1	Full River	1
2	West Tanjung Jabung, Jambi City	2
3	Kerinci, Merangin, Sarolangun, Batanghari, East Tanjung Jabung, Tebo, Bungo	8

From Table 6, we can see the results of grouping districts and cities in Jambi Province into 3 clusters, with the first cluster having 1 member, the second cluster having 2 members, and the third cluster having 8 members.

3.4 Interpretation

According to the data we have collected, here is the interpretation of the results presented in the following form.

Table 7. Interpretation of Three Cluster Grouping

Cluster	V_1	V_2	Average
Cluster 1	0.957	2,321	1,639
Cluster 2	-1,562	0.538	-0.51225
Cluster 3	0.27075	-0.4245	-0.076875

Based on Table 7, it can be interpreted that the districts and cities in Cluster 1 are districts and cities that are at a high level of health. This means that Sungai Penuh City is a city with a high or very good level of health compared to other districts and cities in Jambi Province.

The districts and cities that are members of Cluster 2 are districts and cities that have a moderate level of health compared to other districts and cities. This means that Tanjung Jabung Barat districts and Jambi City are districts and cities that have a moderate or good level of health compared to other districts and cities in Jambi Province.

The districts and cities that are members of Cluster 3 are districts and cities that have low health levels when compared to other districts and cities. This means that Kerinci, Merangin, Sarolangun, Batanghari, East Tanjung Jabung, Tebo, and Bungo districts are districts and cities that have low or poor health levels compared to other districts and cities in Jambi Province.

CONCLUSION

Based on the results of the grouping that has been described in the result and discussion section, it can be concluded that there are 3 levels of clusters in Jambi Province based on its health indicator data. Cluster 1 which contains Sungai Penuh City is at a high level, meaning that the health conditions in the city are very good compared to other cities. Cluster 2 which contains West Tanjung Jabung and Jambi City is at a moderate level, meaning that the health conditions in this city are quite good compared to other cities. Cluster 3 Kerinci, Merangin, Sarolangun, Batanghari, East Tanjung Jabung, Tebo, and Bungo are at a low level, meaning that health conditions in these cities are less good than in other cities..

The researcher recognizes that this study has limitations, which are expected to be developed by future researchers: The use of case studies with data on public health indicators in 11 districts and cities in Jambi Province in 2022, so the findings may not be extrapolated, the data of this study are exclusively health indicators of Jambi Province, and the analysis focuses on health cluster grouping.

REFERENCES

Beattie, J. R., & Esmonde-white, F. W. L. (2021). *Exploration of Principal Component Analysis : Deriving Principal Component Analysis Visually Using Spectra*. *75*(4), 361–375. https://doi.org/10.1177/0003702820987847

Badan Pusat Statistika. (2023). *Profil Statistika Kesehatan 2023*.

BPS. (2022). *Analisis Statistik Kesehatan dan Perumahan Provinsi Jambi 2022*. Jakarta: CV Salim Media Indonesia.

Chen, Y. (2024). Computer-Aided Recognition Algorithm for Students ' Mental Health Problems using K-means Clustering. *Computer-Aided Design and Applications*, *21*(S9), 19–37. https://doi.org/10.14733/cadaps.2024.S9.19-37

Choi, J., Ki, M., Kwon, H. J., Park, B., Bae, S., Oh, C. M., Chun, B. C., Oh, G. J., Lee, Y. H., Lee, T. Y., Cheong, H. K., Choi, B. Y., Park, J. H., & Park, S. K. (2019). Health indicators related to disease, death, and reproduction. *Journal of Preventive Medicine and Public Health*, *52*(1), 14–20. https://doi.org/10.3961/jpmph.18.250

Hafizurrachman. (2020). *Manajemen Pendidikan Dan Kesehatan*.

Hastari, D., Nurunnisa, F., Winanda, S., & Aprillia, D. D. (2023). *Application of K-Means and K-Medoids Algorithms for Grouping Country Data Based on Socio-Economic and Health Factors Penerapan Algoritma K-Means dan K-Medoids untuk Mengelompokkan Data Negara Berdasarkan Faktor Sosial-Ekonomi dan Kesehatan*. 274–281.

Jannati, A. S. R. (2022). Perlindungan Hukum Bagi Pasien Dalam Pelayanan Telemedicine di Indonesia. *Jurnal Juristic*, *3*(02), 210–224.

Japar, M., Semendawai, A. H., Fahruddin, M., Sosial, F. I., Jakarta, U. N., Hukum, F., Jakarta, U. N., & Normatif, P. Y. (2024). *Hukum kesehatan ditinjau dari perlindungan hak asasi manusia*. *5*(1), 952–961.

Mubarak, H., & Kholijah, G. (2023). Analisis Klaster Dalam Pengelompokan Kabupaten/Kota Di Provinsi Jambi Berdasarkan Penyakit Menular Menggunakan Metode K-Means. *Jurnal Statistika Dan Komputasi*, *2*(1), 20–30. https://doi.org/10.32665/statkom.v2i1.1597

Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT*, *4*(1), 20–24. https://doi.org/10.30591/jpit.v4i1.1253

Prakoso, B. H., Rachmawati, E., Mudiono, D. R. P., Vestine, V., & Suyoso, G. E. J. (2023). Klasterisasi Puskesmas dengan K-Means Berdasarkan Data Kualitas Kesehatan Keluarga dan Gizi Masyarakat. *Jurnal Buana Informatika*, *14*(01), 60–68. https://doi.org/10.24002/jbi.v14i01.7105

Ramadhani, D., Setiawan, T. H., Basir, C., & Sari, D. P. (2024). Enhancing Orchid Inventory Management with K-Means Clustering: A Case Study in Sales Optimization. *Mathline : Jurnal Matematika Dan Pendidikan Matematika*, *9*(1), 1–12. https://doi.org/10.31943/mathline.v9i1.529

Ramalisa, Y., Febriyanti, A., & Multahadah, C. (2019). *Analysis of Non Hierarchical Bomb for Collection of Community Health Degrees in Jambi and Muaro Jambi City*. *20*(1), 25–34.

Santoso, S. (2014). Mahir Statistik Multivariat Dengan SPSS. *Jakarta: PT Elek Media Komputindo*.

Saputra. L, Rosalina, R. Oktarina, I. (2023). *Buku Ajar Statistika*. Padang: CV Muharika Rumah Ilmiah.

Sari, Y. R., Sudewa, A., Lestari, D. A., & Jaya, T. I. (2020). Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer. *CESS (Journal of Computer Engineering, System and Science)*, *5*(2), 192. https://doi.org/10.24114/cess.v5i2.18519

Sauddin, I. A. (2021). *Statistika Multivariat*. Alauddin Universitas Press.

Subargus, A. (2022). *Promosi Kesehatan Melalui Pendidikan KESMAS*.

Wardono, W., Sunarmi, S., & Wirawan, M. R. (2019). Pengelompokkan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Dengan Metode K-Means Cluster. *Seminar Nasional Edusaintek*, *2*(1), 599–610. https://prosiding.unimus.ac.id/index.php/edusaintek/article/download/531/534