Journal Health Information Management Indonesian (JHIMI)
ISSN 2829-6435

Analysis of Life Expectancy (LE) in Indonesia Using the K-Means Clustering Algorithm Method for 2020-2023

[1]Deni Luvi Jayanto, [2]Prakasit Poonwong, [3]Vicky Bin Djusmin, [4]Mohamed Fal Mohamed Fadel
[5]Nimatu Zuliana, [6]Ninda Mulya Ike Ardila, [7]Mia Ashari Kurniasari
[1] Universitas Tadulako, Sulawesi Tengah
[2]Praboromrajchanok Institute Thailand
[3]Universitas Cokroaminoto Palopo, Sulawesi Selatan
[4]Al Akhawayn University, Morocco
[5,6,7]Institut Ilmu Kesehatan Bhakti Wiyata
[1]Jl. Soekarno Hatta KM. 9, Tondo, Mantikulore, Palu, Sulawesi Tengah 94148, Indonesia
[2]Building 4, 8th Floor, Office of the Permanent Secretary, Ministry of Public Health, Tiwanon Rd, Subdistrict, Mueang Nonthaburi District, Nonthaburi 11000, Thailand
[3]Jl. Latamacelling, Tompotika, Wara, Palopo, Sulawesi Selatan 91911, Indonesia
[4]Hassan II Avenue, PO Box 104. Ifrane 53000 Fes-Meknes, Morocco
[5,6,7]Jl. KH Wachid Hasyim No.65, Kediri, Kota Kediri, Jawa Timur 64114, Indonesia
* deniluvi@untad.ac.id

Uploaded: 2025-02-03, Revised: 2025-08-03, Accepted: 2025-12-10

Abstract — Life Expectancy (LE) is one of the key indicators in assessing the quality of health in a region. This study aims to analyze the distribution patterns of LE in Indonesia from 2020 to 2023 using the K-Means Clustering algorithm. The data used includes LE from all provinces in Indonesia during this period, sourced from secondary data provided by the Central Bureau of Statistics (BPS). This research adopts a quantitative descriptive approach with a population comprising all provinces in Indonesia for the years 2020-2023. The analysis results indicate that provinces can be grouped into three clusters based on LE levels: high LE cluster (C1), moderate LE cluster (C2), and low LE cluster (C3). High LE clusters, such as DKI Jakarta and Bali, demonstrate good healthcare infrastructure, high education levels, and effective government programs. Conversely, low LE clusters, such as Central Kalimantan and North Sumatra, face challenges such as limited access to healthcare services, low education levels, and high poverty rates. This study recommends the development of healthcare infrastructure, equitable distribution of healthcare resources, and the enhancement of health education programs in low LE regions to reduce disparities across regions.

Keywords – Life Expectancy, K-Means Clustering, Health Disparities, Healthcare Infrastructure, Indonesia

Abstrak — Angka Harapan Hidup (AHH) merupakan salah satu indikator penting dalam menilai kualitas kesehatan suatu wilayah. Penelitian ini bertujuan untuk menganalisis pola distribusi AHH di Indonesia pada tahun 2021-2023 menggunakan algoritma K-Means Clustering. Data yang digunakan mencakup AHH dari seluruh provinsi di Indonesia selama periode tersebut yang diambil dari data sekunder di Badan Pusat Statistik (BPS). Jenis penelitian menggunakan deskriptif kuantitatif dengan data populasi semua provinsi yang ada di Indonesia pada tahun 2021-2023. Hasil analisis menunjukkan bahwa provinsi-provinsi dapat dikelompokkan ke dalam tiga klaster berdasarkan tingkat AHH, yaitu klaster dengan AHH tinggi (C1), sedang (C2), dan rendah (C3). Klaster dengan AHH tinggi, seperti DKI Jakarta dan Bali, menunjukkan infrastruktur kesehatan yang baik, tingkat pendidikan yang tinggi, serta efektivitas program pemerintah. Sebaliknya, klaster dengan AHH rendah, seperti Kalimantan Tengah dan Sumatera Utara, menghadapi tantangan berupa keterbatasan akses layanan kesehatan, tingkat pendidikan rendah, dan tingginya tingkat kemiskinan. Penelitian ini merekomendasikan pengembangan infrastruktur kesehatan, pemerataan sumber daya kesehatan, serta peningkatan program pendidikan kesehatan di wilayah dengan AHH rendah untuk mengurangi kesenjangan antar wilayah.

Kata kunci – Angka Harapan Hidup, K-Means Clustering, Disparitas Kesehatan, Infrastruktur Kesehatan, Indonesia

Journal Health Information Management Indonesian (JHIMI)
ISSN 2829-6435

1. INTRODUCTION

Life expectancy or *Angka Harapan Hidup (AHH)* is one of the main indicators used to measure the quality of life and well-being of a society. This indicator reflects the level of health, access to healthcare services, and the socio-economic conditions of a region. According to a report from Indonesia's Central Bureau of Statistics (BPS), life expectancy in Indonesia has steadily increased over the past few decades. However, significant disparities between provinces remain [1]. Although the national life expectancy has improved, regional disparities continue to pose a challenge. Provinces in Java and Bali tend to have higher life expectancy compared to regions outside Java, such as Nusa Tenggara, Maluku, and Papua. This highlights inequalities in access to healthcare services, infrastructure, and education [2]. Life expectancy is influenced by various factors, such as poverty levels, access to healthcare services, education, and infrastructure. Previous studies have shown that regions with higher education levels and adequate healthcare access tend to have better life expectancy [3]. Therefore, data analysis is necessary to understand the patterns and factors influencing these disparities.

The K-Means clustering method has been widely used in social data analysis to group regions based on certain characteristics. In the context of life expectancy, this method can help identify patterns and group provinces with similar characteristics, thereby facilitating data-driven decision-making [4]. By clustering provinces based on life expectancy, the government can more easily determine intervention priorities. Provinces in clusters with low life expectancy, for instance, can become the main focus of poverty alleviation programs, improved access to healthcare services, and public education initiatives [5].

K-Means clustering works by dividing data into several clusters based on the minimum distance to the cluster center (centroid). This method has been widely applied in various studies, including in the field of public health. For example, a study demonstrated that K-Means can be used to group regions based on healthcare accessibility levels, ultimately assisting governments in prioritizing the development of healthcare infrastructure [6].

In the context of life expectancy, clustering becomes an important tool for understanding regional disparities. Previous studies have shown that regions with low life expectancy often share similar characteristics, such as high poverty rates, low education levels, and limited access to healthcare services [2]. By clustering these regions, the government can design more specific and effective interventions.

The advantage of the K-Means clustering method lies in its ability to handle large datasets with high efficiency. Additionally, clustering provides deeper insights into regional patterns that may not be visible in simple descriptive analyses. For instance, research conducted showed that clustering can identify regions with urgent healthcare needs, which may not be revealed by merely observing average life expectancy [7].

This study not only provides an overview of life expectancy disparities in Indonesia but also demonstrates the potential use of clustering in data-driven policy planning. By understanding the existing patterns, the government can allocate resources more effectively to address regional inequalities. Clustering can also be used to monitor changes in patterns over time, enabling better policy evaluation [5]

Thus, the use of clustering, particularly the K-Means method, in life expectancy research in Indonesia contributes significantly to supporting more inclusive and sustainable development planning. This analysis not only helps identify groups of provinces with similar characteristics but also provides a strong foundation for designing interventions tailored to the needs of each region.

Unequal healthcare infrastructure is one of the main causes of life expectancy disparities. Urban areas tend to have better healthcare facilities compared to remote regions. This disparity affects people's ability to access quality healthcare services [2]. Education plays an important role in improving life expectancy because individuals with higher education levels tend to have greater awareness of health and healthy lifestyles. Provinces with low education levels often fall into groups with low life expectancy [8].

Socio-economic factors, such as income levels and poverty, also contribute to differences in life expectancy. Provinces with high poverty

rates, such as those in East Nusa Tenggara and Papua, tend to have lower life expectancy compared to other regions [1]. Data-driven analyses, such as K-Means clustering, are essential to support more targeted policies. By understanding the existing patterns, the government can design more effective and efficient programs to improve the quality of life across all provinces [9]

This study aims to provide a clearer picture of life expectancy disparities in Indonesia through clustering analysis. The results are expected to serve as a reference for policymakers in designing interventions tailored to the needs of each region, thereby reducing inequalities and improving societal well-being more equitably [4].

2. RESEARCH METHODS

This study aims to analyze the patterns of life expectancy across various provinces in Indonesia using the K-Means clustering method. The research methodology is systematically designed to ensure valid and reliable results. The following are the stages of the methodology:

1. Type of Research

This research is a descriptive quantitative study with a secondary data analysis approach. The data used consists of life expectancy figures from 2020 to 2023 obtained from reliable sources, such as the Central Bureau of Statistics (BPS).

2. Data Sources

The data used in this study is secondary data, which includes life expectancy figures from all provinces in Indonesia. The data sources include Central Bureau of Statistics (BPS) for Annual life expectancy data for each province.

3. Data Collection Method

Data collection was conducted by accessing official publications from BPS.

4. Data Analysis Stages

This study utilizes the K-Means clustering method to analyze the data. The stages of data analysis include:

a. Data Preprocessing, Cleaning the data to ensure there are no missing or invalid values.
b. Normalizing the data to ensure all variables are on the same scale, preventing bias in the clustering results.
c. Determining the Number of Clusters, Identifying the optimal number of clusters using the Elbow Method, which evaluates changes in inertia values (within-cluster sum of squares) across different numbers of clusters. The selected number of clusters corresponds to the point where the reduction in inertia begins to slow down.
d. K-Means Implementation, Applying the K-Means algorithm to group provinces based on life expectancy. The algorithm works by randomly determining initial centroids, calculating the distance of data points to the centroids, and grouping data into clusters based on the minimum distance. This process is repeated until convergence is achieved.
e. Result Interpretation, Grouping provinces into clusters based on life expectancy patterns.

5. Software Used

The following software is used for data analysis:

a. Weka, For data preprocessing, K-Means algorithm implementation, and clustering visualization.
b. Microsoft Excel, For initial data processing.

6. Analytical Framework

The analytical framework of this study is based on a clustering-based data analysis approach, consisting of three main stages:

a. Descriptive Analysis: Presenting life expectancy data as a whole to provide an initial overview.
b. K-Means Clustering: Grouping provinces based on life expectancy patterns.
c. Clustering Result Analysis: Identifying the characteristics of each cluster and providing policy recommendations based on the analysis results.

3. RESULTS

In the process of testing the K-Means Clustering method in a data grouping system, data is required as input to support the analysis process. Based on the review of BPS data from 2020-2023, the collected data will be used as the primary material for analysis in this study using the K-Means Clustering method.

1. Data Preprocessing
The preprocessing stage has been conducted, which involves data cleaning. During this stage, it was found that some data was removed due to empty values. The data that was removed includes the provinces of Papua Barat Daya, Papua Selatan, Papua Tengah, Papua Pegunungan. Subsequently, the file was converted from the xlsx extension to csv so that it could be read by the processing application.

2. Data Normalizing
   In this stage, the researcher performed data normalization by re-examining the existing data structure to ensure a consistent scale. It was found that there were differences in the use of commas in numbers that should have used periods. The data is detailed as follows **Table 2**.
3. Determining the Number of Clusters
   The clustering stage has been performed with the Within Cluster Sum of Squared Errors using the elbow method as follows.

Table 1. Sum of squared errors and cenroid clusters Year 2020-2023 (n=4)

Year	Within cluster sum of squared errors and cenroid clusters.	
	Male	**Female**
2020	31.320982998 Centroid Cluster 0: 71.2389 Centroid Cluster 1: 64.4943 Centroid Cluster 2: 68.0728	31.321397733 Centroid Cluster 0: 74.6791 Centroid Cluster 1: 67.975 Centroid Cluster 2: 71.6653
2021	31.3255850863 Centroid Cluster 0: 71.36 Centroid Cluster 1: 64.6371 Centroid Cluster 2: 68.1683	31.336710941 Centroid Cluster 0: 74.7845 Centroid Cluster 1: 68.1417 Centroid Cluster 2: 71.7712
2022	31.3308121476 Centroid Cluster 0: 71.51 Centroid Cluster 1: 64.9129 Centroid Cluster 2: 68.4033	31.330827079 Centroid Cluster 0: 75.14 Centroid Cluster 1: 68.5217 Centroid Cluster 2: 72.1517
2023	31.3346095183 Centroid Cluster 0: 71.579 Centroid Cluster 1: 64.9583 Centroid Cluster 2: 68.5067	31.320234181 Centroid Cluster 0: 75.48 Centroid Cluster 1: 68.8933 Centroid Cluster 2: 72.4617

Table 1, contains of cluster sum of squared errors and centroid clusters are presented for each year and for males and females. These centroid values are used as the central point values for calculating the proximity of data to belong to cluster 1, cluster 2, and cluster 3.

4. K-Means Implementation
   Cluster calculations have been performed based on the centroids that have been found. The results of the K-Means cluster can be seen in **Table 3**
5. Result Interpretation
   Table 2: In the table above, the trend of Life Expectancy (LE) in Indonesia shows an increase in almost all provinces during the period from 2020 to 2023. DI Yogyakarta recorded the highest LE in 2023, with 73.40 years for males and 77.03 years for females. In addition to Yogyakarta, other provinces with high LE include DKI Jakarta, Central Java, and Bali. Conversely, West Sulawesi had the lowest LE in 2023, at 64.11 years for males and 68.00 years for females, followed by Papua, West Papua, and Maluku.

In general, females have a higher LE compared to males in all provinces, with an average difference of around 4 to 5 years. In terms of changes, several provinces experienced significant increases in LE. For instance, West Nusa Tenggara (NTB) saw an increase from 64.63 years in 2020 to 65.54 years in 2023 for males. Gorontalo also showed an increase from 66.14 years to 66.90 years over the same period, as did West Sulawesi, which rose from 63.20 years to 64.11 years.

Table 3: This data reflects improvements in the health and well-being sectors across various regions in Indonesia. Based on the clustering analysis using the K-Means method, provinces in Indonesia are grouped into three clusters (C1, C2, and C3) based on life expectancy from 2020 to 2023. Cluster C1 includes provinces with relatively high life expectancy, such as DKI Jakarta, West Java, Central Java, DI Yogyakarta, and Bali. These provinces have good healthcare facilities, adequate access to health services, as well as high levels of education and public awareness. Cluster C2 consists of provinces with moderate life expectancy, such as West Nusa Tenggara, East Nusa Tenggara, Maluku, and West Papua. These regions still face challenges in access to health services, infrastructure, or other socio-economic factors. Meanwhile, Cluster C3 includes provinces with relatively low life expectancy, such as North Sumatra, South Sumatra, Bengkulu, and Central Kalimantan, which are likely to face obstacles such as limited access to healthcare facilities, low education levels, and higher poverty rates.

The observed pattern shows that provinces on Java Island and Bali dominate the cluster with high life expectancy, which can be attributed to urbanization, adequate health infrastructure, and more focused government programs. Conversely, regions in Clusters C2 and C3 reflect disparities in life expectancy across various areas of Indonesia. To address this, data-driven interventions are needed, such as the development of basic health services, improved accessibility to healthcare facilities, and health education programs, especially in Cluster C3 regions. In addition, Cluster C2 requires strengthened infrastructure and equitable healthcare services to improve the quality of life of its population. Provinces in Cluster C1 can serve as models for developing effective health policies in other regions. This analysis highlights the importance of government efforts in reducing

disparities in life expectancy between regions in Indonesia.

Table 2. Data on Life Expectancy (LE) by Province in Indonesia for the Years 2020-2023 (n = 34)

Provinsi	2020		2021		2022		2023	
	Male	Female	Male	Female	Male	Female	Male	Female
Aceh	68.04	71.91	68.07	71.94	68.26	72.16	68.42	72.35
Sumatera Utara	67.22	71.08	67.35	71.21	67.66	71.6	68.06	72
Sumatera Barat	67.59	71.45	67.7	71.57	67.99	71.89	68.27	72.21
Riau	69.75	73.55	69.82	73.62	70.03	73.9	70.28	74.29
Jambi	69.27	73.07	69.33	73.19	69.57	73.49	69.83	73.81
Sumatera Selatan	68	71.86	68.11	71.95	68.38	72.29	68.74	72.68
Bengkulu	67.47	71.27	67.54	71.4	67.74	71.68	67.99	71.94
Lampung	68.78	72.61	68.86	72.69	69.07	72.97	69.31	73.28
Kep. Bangka Belitung	68.77	72.59	68.86	72.7	69.06	72.97	69.3	73.26
Kep. Riau	68.08	71.9	68.24	72.07	68.58	72.47	68.99	72.94
DKI Jakarta	71.1	74.8	71.21	74.91	71.45	75.22	71.88	75.61
Jawa Barat	71.3	75	71.57	75.19	71.77	75.48	72.01	76.19
Jawa Tengah	72.51	76.3	72.61	76.42	72.71	76.53	72.95	76.62
DI Yogyakarta	73.22	76.83	73.27	76.89	73.28	76.93	73.4	77.03
Jawa Timur	69.42	73.27	69.51	73.35	69.81	73.71	70.16	74.16
Banten	68.08	71.93	68.15	71.99	68.46	72.36	68.85	72.79
Bali	70.28	74.03	70.39	74.17	70.69	74.53	70.99	75.07
Nusa Tenggara Barat	64.63	68.39	64.81	68.66	65.14	69.07	65.54	69.5
Nusa Tenggara Timur	65.15	68.96	65.28	69.11	65.54	69.43	65.82	69.8
Kalimantan Barat	68.8	72.67	68.87	72.74	69.08	73	69.38	73.35
Kalimantan Tengah	67.84	71.66	67.89	71.72	68.08	72.02	68.36	72.29
Kalimantan Selatan	66.76	70.62	66.97	70.79	67.23	71.13	67.49	71.45
Kalimantan Timur	72.54	76.21	72.79	76.51	72.8	76.52	72.96	76.62
Kalimantan Utara	70.63	74.44	70.69	74.48	70.7	74.54	70.72	74.75
Sulawesi Utara	69.82	73.67	69.89	73.72	70.16	74.04	70.44	74.46
Sulawesi Tengah	66.77	70.72	66.91	70.85	66.98	70.95	67.24	71.2
Sulawesi Selatan	68.68	72.56	68.77	72.65	69.03	72.96	69.28	73.24
Sulawesi Tenggara	69.31	73.37	69.35	73.37	69.44	73.47	69.54	73.51
Gorontalo	66.14	70.09	66.27	70.21	66.56	70.53	66.9	70.87
Sulawesi Barat	63.2	67.02	63.39	67.19	63.74	67.6	64.11	68
Maluku	64.12	67.93	64.23	68.05	64.53	68.43	64.87	68.8
Maluku Utara	66.43	70.32	66.55	70.44	66.84	70.79	67.17	71.14
Papua Barat	64.2	67.9	64.33	68.05	64.54	68.44	64.88	68.81
Papua	64.02	67.65	64.15	67.79	64.34	68.16	64.53	68.45

Table 2 . Data from K-Means Cluster Results on Life Expectancy (LE) by Province in Indonesia for the Years 2020-2023 (n = 34)

Province	2020		2021		2022		2023	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Aceh	C3	C3	C3	C3	C3	C3	C3	C3
Sumatera Utara	C3	C3	C3	C3	C3	C3	C3	C3
Sumatera Barat	C3	C3	C3	C3	C3	C3	C3	C3
Riau	C1	C1	C1	C1	C1	C1	C1	C1
Jambi	C3	C3	C3	C3	C3	C3	C3	C3
Sumatera Selatan	C3	C3	C3	C3	C3	C3	C3	C3
Bengkulu	C3	C3	C3	C3	C3	C3	C3	C3
Lampung	C3	C3	C3	C3	C3	C3	C3	C3
Kep. Bangka Belitung	C3	C3	C3	C3	C3	C3	C3	C3
Kep. Riau	C3	C3	C3	C3	C3	C3	C3	C3
DKI Jakarta	C1	C1	C1	C1	C1	C1	C1	C1
Jawa Barat	C1	C1	C1	C1	C1	C1	C1	C1
Jawa Tengah	C1	C1	C1	C1	C1	C1	C1	C1
DI Yogyakarta	C1	C1	C1	C1	C1	C1	C1	C1
Jawa Timur	**C3**	**C1**	**C3**	**C1**	**C3**	**C1**	**C1**	**C1**
Banten	C3	C3	C3	C3	C3	C3	C3	C3
Bali	C1	C1	C1	C1	C1	C1	C1	C1
Nusa Tenggara Barat	C2	C2	C2	C2	C2	C2	C2	C2
Nusa Tenggara Timur	C2	C2	C2	C2	C2	C2	C2	C2
Kalimantan Barat	C3	C3	C3	C3	C3	C3	C3	C3
Kalimantan Tengah	C3	C3	C3	C3	C3	C3	C3	C3
Kalimantan Selatan	C3	C3	C3	C3	C3	C3	C3	C3
Kalimantan Timur	C1	C1	C1	C1	C1	C1	C1	C1
Kalimantan Utara	C1	C1	C1	C1	C1	C1	C1	C1
Sulawesi Utara	C1	C1	C1	C1	C1	C1	C1	C1
Sulawesi Tengah	C3	C3	C3	C3	C3	C3	C3	C3
Sulawesi Selatan	C3	C3	C3	C3	C3	C3	C3	C3
Sulawesi Tenggara	**C3**	**C1**	**C3**	**C1**	**C3**	**C3**	**C3**	**C3**
Gorontalo	**C2**	**C3**	**C2**	**C3**	**C2**	**C3**	**C3**	**C3**
Sulawesi Barat	C2	C2	C2	C2	C2	C2	C2	C2
Maluku	C2	C2	C2	C2	C2	C2	C2	C2
Maluku Utara	C3	C3	C3	C3	C3	C3	C3	C3
Papua Barat	C2	C2	C2	C2	C2	C2	C2	C2
Papua	C2	C2	C2	C2	C2	C2	C2	C2

4. DISCUSSION

The trend of Life Expectancy (LE) in Indonesia shows a significant increase during the period from 2020 to 2023. This reflects improvements in the health sector, access to health services, and community welfare programs. DI Yogyakarta is recorded as the province with the highest LE in 2023, at 73.40 years for males and 77.03 years for females. This aligns with research indicating that factors such as education and better access to health care in this region contribute to higher life expectancy [10].

Conversely, West Sulawesi has the lowest LE, at 64.11 years for males and 68.00 years for females. Research shows that areas with inadequate health infrastructure tend to have lower LE [11]. In general, females have a higher LE compared to males, with an average difference of about 4 to 5 years. This is consistent with global findings indicating that females tend to live longer than males [12]. Several provinces have experienced a significant increase in LE, such as West Nusa Tenggara (NTB) and Gorontalo. This increase indicates efforts by the government and society to improve the quality of life and health [1].

This data reflects improvements in the health sector and community welfare. Investments in health services, education, and quality of life enhancement programs are key factors in achieving these results [13]. Overall, the trend of LE in Indonesia shows positive progress, but there are still challenges in certain provinces that require further attention. Ongoing efforts to improve health services and community welfare are essential to achieving better development targets in the future [14].

The COVID-19 pandemic had a significant impact on Life Expectancy (LE). Although there was an increase before the pandemic, many reports indicate that the pandemic caused a temporary decline in LE due to rising mortality rates and disruptions in health services [12]. Government policies focused on improving access to health services, such as the National Health Insurance (JKN), have helped increase LE. These programs aim to ensure that all citizens have access to adequate health services [14].

Socioeconomic factors also play a crucial role in determining Life Expectancy (LE) [15]. Research shows that communities with higher levels of education and better income tend to have higher LE. Compared to other countries in Southeast Asia, Indonesia's LE shows progress, although it still lags behind some neighboring countries such as Malaysia and Thailand [16]. Despite the improvements, there are still challenges to be faced, such as health disparities between provinces and access to health services in remote areas [17].

Based on these trends and analyses, several policy recommendations can be proposed:

1. Improving Healthcare Access in Underserved Areas, The government needs to expand access to healthcare facilities in regions with low LE, such as Papua, West Papua, and Maluku.
2. Strengthening Disease Prevention Programs, Programs such as immunization, control of infectious diseases, and stunting prevention need to be continuously strengthened to improve the quality of life of the population.
3. Increasing Investment in the Education Sector, Given the close relationship between education and health, the government needs to improve access to education in regions with low LE.

Analysis results of clustering using the K-Means method show a clear pattern in the distribution of life expectancy in Indonesia. Provinces are grouped into three clusters based on life expectancy from 2020 to 2023, reflecting disparities in access to healthcare services, infrastructure, and socio-economic conditions across regions. Urbanization in these areas has driven the development of better infrastructure, including hospitals, community health centers, and other healthcare facilities. According to the theory, urbanization is often positively correlated with increased access to basic services, including healthcare [18].

1. Cluster C1: Provinces with High Life Expectancy

Higher education levels in these regions contribute to public awareness of the importance of healthy lifestyles and the use of healthcare services. Research by Grossman (1972) in "The Demand for Health" highlights that education plays a crucial role in shaping individual health behavior [19]. These areas are often prioritized in government programs, such as the National Health Insurance (JKN) and immunization initiatives. A study by Nugroho (2020) in the Indonesian Health Policy journal indicates that the implementation of health programs in urban areas tends to be more effective due to better accessibility [20].

2. Cluster C2: Provinces with Moderate Life Expectancy

Provinces such as West Nusa Tenggara, East Nusa Tenggara, Maluku, and West Papua fall into this cluster. These regions still face significant challenges, such as:

a. Infrastructure Limitations, Healthcare infrastructure in these areas remains inadequate, especially in remote regions. According to a report by Bappenas (2021), many areas in Cluster C2 have doctor-to-population ratios and healthcare facilities far below national standards [21].
b. Socio-Economic Factors, Higher poverty rates and limited access to education also influence life expectancy. Research by Marmot (2005) in The Social Determinants of Health emphasizes that socio-economic factors have a significant impact on public health [22].
c. Inequities in Healthcare Service Distribution, Uneven distribution of medical personnel and healthcare services remains a major challenge. A

study by Kusnandar (2020) in the journal Health Equity in Indonesia shows that eastern Indonesia tends to have more limited access to healthcare services compared to western regions [23].

3. Cluster C3: Provinces with Low Life Expectancy

Provinces such as North Sumatra, South Sumatra, Bengkulu, and Central Kalimantan are included in this cluster. Several factors contribute to the low life expectancy in these areas, including:

a. Lack of Access to Healthcare Facilities, These regions face limitations in both the quantity and quality of healthcare facilities. According to the WHO (2022) report, physical access to healthcare facilities in remote areas is often a major obstacle [12].
b. Low Education and High Poverty Rates, Low education levels and higher poverty rates in these regions reduce the population's ability to maintain their health. Research by Luy and Gast (2014) highlights the close relationship between education, income, and health [24].
c. Minimal Government Intervention, Health programs in these areas are often less effective due to limited resource allocation and geographical challenges. A study by Bappenas (2021) underscores the need for data-driven approaches to enhance the effectiveness of health programs in these regions [21].
d. Patterns and Policy Implications, the pattern observed in this analysis shows that provinces in Java and Bali dominate the cluster with high life expectancy, reflecting the success of urbanization and healthcare infrastructure development in these regions. Conversely, regions in Clusters C2 and C3 reflect significant disparities in life expectancy.

To address these disparities, the following steps are needed:

a. Development of Basic Healthcare Services, Regions in Cluster C3 require the development of basic healthcare facilities, such as community health centers and clinics, to improve accessibility to healthcare services.
b. Enhancement of Health Education Programs, Health education programs focusing on disease prevention and healthy lifestyles need to be improved, especially in Clusters C2 and C3.
c. Equitable Healthcare Infrastructure Distribution, The government needs to strengthen healthcare infrastructure in Cluster C2 regions, including the distribution of medical personnel and healthcare facilities.
d. Using Cluster C1 as a Model, Provinces in Cluster C1 can serve as a model for developing effective health policies in other regions.
e. By implementing these strategic measures, the government is expected to reduce disparities in life expectancy across regions and improve the overall quality of life for the Indonesian population.

5. CLOSING

Conclusion

The clustering analysis using the K-Means method reveals significant disparities in life expectancy across Indonesian provinces, grouped into three distinct clusters. Cluster C1, which includes provinces such as DKI Jakarta, Bali, and Yogyakarta, exhibits high life expectancy due to better healthcare infrastructure, higher education levels, and effective government programs. Cluster C2, comprising provinces like West Nusa Tenggara and Maluku, faces moderate life expectancy challenges due to limited healthcare infrastructure, socio-economic constraints, and uneven distribution of healthcare services. Meanwhile, Cluster C3, which includes provinces like North Sumatra and Central Kalimantan, has low life expectancy caused by inadequate healthcare facilities, low education levels, high poverty rates, and minimal government intervention. These findings highlight the urgent need to address the inequalities in healthcare access and socio-economic conditions across regions to improve overall life expectancy.

Suggestion

Prioritize the development of healthcare facilities, such as community health centers and clinics, in provinces within Cluster C3 to ensure better access to basic health services. Implement comprehensive health education programs focusing on disease prevention, healthy lifestyles, and the importance of utilizing healthcare services, particularly in Clusters C2 and C3. Improve the allocation and distribution of medical personnel and healthcare infrastructure in Cluster C2 regions to reduce disparities in healthcare access. Use provinces in Cluster C1 as a model for developing effective health policies and infrastructure in other regions, adapting strategies to the specific needs of Clusters C2 and C3. By implementing these recommendations, the government can work toward reducing disparities, improving life expectancy, and achieving more equitable health outcomes across Indonesia.

6. THANK-YOU NOTE

We sincerely thank all individuals and institutions who contributed to the success of this research through their support, insights, and resources.

7. BIBLIOGRAPHY

[1] Badan Pusat Statistik Indonesia, *Statistik Indonesia 2023*, vol. 1101001. Jakarta: Badan Pusat Statistik, 2023.
[2] Kemenkes RI, “Laporan Tahunan Kesehatan Nasional. Jakarta: Kemenkes RI.,” Jakarta, 2022.
[3] World Health Organization, “World Health Organization Global Health Observatory: Data Repository,” World Health Organization.

[4] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and TechniquesHan, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. San Francisco, CA, itd: Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-381479-1.00001-0*. 2012.
[5] Bappenas, "Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024," Jakarta, 2023.
[6] M. , C. M. , & I. M. Rahman, "Clustering-Based Analysis of Healthcare Access in Developing Countries," *Int J Public Health Res*, vol. 45, no. 2, pp. 123–135, 2021.
[7] Q. , Z. X. , & L. Y. Wu, "Regional Health Disparities: A Clustering Approach to Identify Priority Areas," *Journal of Health Geography*, vol. 18, no. 4, pp. 89–102, 2020.
[8] UNESCO, "Education for Sustainable Development," Paris, 2020.
[9] McKinsey & Company, *The Role of Data in Public Policy*. New York: McKinsey, 2021.
[10] A. Susilo, *Health Policy and Management in Indonesia*. Jakarta: Universitas Indonesia, 2021.
[11] D. Rachmawati, *Public Health Challenges in Indonesia*. Yogyakarta: Graha Ilmu, 2020.
[12] WHO, *World health statistics 2022 (Monitoring health of the SDGs)*. 2022.
[13] R. Widiastuti, *Health and Development in Indonesia: A Review*. Jakarta: Penerbit Erlangga, 2023.
[14] Kemenkes RI, "Laporan Tahunan Kesehatan 2023," Jakarta, 2023.
[15] B. Setiawan, *Socioeconomic Factors and Life Expectancy in Indonesia*. Jakarta: Salemba Empat, 2022.
[16] Asean Secretariat, "Asean Matters: Epicentrum Of Growth," Jakarta, 2022.
[17] N. Nursalam, *Challenges in Public Health in Indonesia*. Yogyakarta: Andi Offset, 2021.
[18] M. P. Todaro and S. C. Smith, *Economic Development*, 12th ed. Pearson Education, 2015.
[19] M. Grossman, *The Demand for Health*. 2017. doi: 10.7312/gros17900.
[20] S. Nugroho, "The Effectiveness of Health Programs in Urban Areas," *Indonesian Health Policy*, vol. 12, no. 1, pp. 45–58, 2020.
[21] Bappenas, "Laporan Kesehatan dan Pembangunan Infrastruktur Kesehatan di Indonesia," Jakarta, 2021.
[22] M. Marmot, "The Social Determinants of Health," *The Lancet*, pp. 1099–1104, 2005.
[23] A. Kusnandar, "Healthcare Access Disparities in Eastern Indonesia," *Health Equity in Indonesia*, vol. 5, no. 3, pp. 210–218, 2020.
[24] M. Luy and K. Gast, "The Relationship between Education, Income, and Health: A Global Perspective," *nternational Journal of Public Health*, vol. 59, no. 2, pp. 235–243, 2014.