Clustering Of Districts Based On Infrastructure Indicators Using K-Means And Average Linkage Methods

Dewi Eliza Agustia[1] ,Noer Fadhly[2*], Cut Zukhrina Oktaviani[3]

[1,2,3] Department of Civil Engineering, Faculty of Engineering, Syiah Kuala University, Banda Aceh
*Coresponding Author:
Email: noer6637@usk.ac.id

Abstract.

Aceh Jaya has great potential in various sectors, such as agriculture, plantations, and tourism. Badan Perencanaan Pembangunan Riset dan Teknologi Daerah (BAPPERIDA) Aceh Jaya District stated that the unequal distribution of infrastructure development in Aceh Jaya District in 2023-2026 resulted in low investment competitiveness and decreased economic performance. Because of these conditions, this study clusters the regions in Aceh Jaya District based on infrastructure indicators to obtain groups of sub-districts based on their level of infrastructure development. The method used in this study is clustering analysis by comparing it with two clustering methods, namely K-Means and Average Linkage, which are validated based on the Silhouette Coefficient value to see which cluster is the best. The data used are infrastructure indicators contained in the Recana Pembangunan Kabupaten Aceh Jaya 2023-2026, namely irrigation area, houses with access to drinking water, sanitised households, drainage, district road length, livable houses, telecommunication services, electricity services. The results obtained are from eight research variables for cluster analysis assumptions with the Kaiser-Mayer-Olkin (KMO) test is 0.611 only seven variables that indicate that the sample is representative, for the zscore value there is no outliar data, which means that the data of the seven infrastructure variables can be used entirely. The clustering results using the K-Means method resulted in 3 clusters and with the Average Lingkage method resulted in 2 clusters, this means that Aceh Jaya District can be clustered into 2 clusters seen from the Silhouette Coefficient value of 0.736. The conclusion that can be drawn from this research is that infrastructure development has not been evenly distributed in the 9 existing sub-districts, so that strategies and priorities are needed in infrastructure development in Aceh Jaya Regency.

Keywords: *Klasterisasi, indikator infrastruktur, and silhouette coefficient.*

I. INTRODUCTION

Infrastructure is the main foundation for economic development and community welfare, as seen from the economic growth of a region. The availability and quality of adequate infrastructure determine the level of accessibility, transportation efficiency, and economic activity of a region. In addition, in the SDGs (Sustainable Development Goals), there are several goals that are directly related to infrastructure, including the 2030 SDG's target of ensuring access for all to decent, safe, affordable housing and basic service infrastructure. The goal highlights the importance of strong, sustainable, and inclusive infrastructure to support economic development, improve quality of life, and promote environmental sustainability. Quality infrastructure stimulates the movement of goods and services and human resources more smoothly and efficiently, thereby increasing economic growth and a better quality of life for the community. This is because infrastructure development is also one of the investments that can increase regional income. The amount of infrastructure investment spent by local governments has a strong influence on the development of economic growth (Masdi & Iskandar, 2021).Aceh Jaya is a district in Aceh Province with significant potential across many industries, including agriculture, plantations, and tourism (Bahri et al., 2021; Puspitaa et al., 2022). Aceh Jaya is also an area that has roads that are very important in improving the economy in the western and southern regions (Isya et al., 2021).

Based on Badan Perencanaan Pembangunan Riset dan Teknologi Daerah (BAPPERIDA) Aceh Jaya District in Rencana Pembangunan Kabupaten (RPK) Aceh Jaya tahun 2023-2026 stated that the unequal distribution of infrastructure development in Aceh Jaya Regency has resulted in low investment competitiveness and declining economic performance. Essential infrastructure still required include the building of habitable residences, flood mitigation, potable water supply, sanitation, waste management, development and enhancement of irrigation systems, drainage, and environmental management and organization. The variations in infrastructure availability and quality can impede regional development and

result in socio-economic inequities. A comprehensive study of the infrastructural situation in each sub-district in Aceh Jaya is necessary to resolve this issue. Clustering analysis serves as an excellent method for categorizing sub-districts according to similarities in infrastructure indicators and aids in identifying clusters of sub-districts with analogous infrastructure features. This information will be important for local government in formulating and executing more focused and efficient development policies.

II. LITERATURE STUDY

Clustering

Clustering is the methodology of categorizing a collection of data items into many groups or clusters, ensuring that things within a cluster exhibit high similarity while being distinctly dissimilar to objects in other clusters. Differences and similarities are evaluated according to attribute values that characterize the objects and frequently entail distance measurements. Clustering is an unsupervised data mining technique, indicating that it is utilized without training, without supervision, and does not necessitate a goal result (Badruddin et al., 2022).Data Mining Concepts and Techniques delineates key clustering needs, specifically scalability and the capacity to manage diverse data kinds. Identification of clusters exhibiting anomalous geometries; Requirements for domain expertise to ascertain parameters; Capability to manage noisy data; Sensitivity to variations in input; Capability to cluster high-dimensional data; constraint-driven clustering; interpretability and usability (Jiawei Han, 2012)

Region

Law No. 26/2007 on Spatial Planning defines a region as an area characterized by geographical challenges and several connected features, with borders and systems established based on administrative or functional criteria. A region is an area characterized by authority, governance, and oversight, encompassing regional environments such as provinces, districts, and subdistricts. The regional environment refers to a defined area, including provinces, districts, and cities (Tambunan et al., 2024). According to the National Spatial Coordinating Agency (2002), a region is characterized as a geographical unit encompassing all associated elements, with boundaries and systems established based on administrative or functional criteria (Bahri et al., 2024).

Infrastructure

(Sanjaya, 2018) Infrastructure is a crucial element of the city encompassing primary construction operations and ancillary structures. Infrastructure systems that deliver transportation, water, structures, and other public amenities essential for fulfilling economic and social requirements of fundamental human needs. The Presidential Regulation of the Republic of Indonesia Number 38 of 2015 defines infrastructure as the technical, physical, systemic, hardware, and software facilities required to deliver services to the community and to support the structural network, thereby facilitating effective economic and social growth. The Presidential Regulation of the Republic of Indonesia number 67 of 2005 delineates the categories of infrastructure as follows: transportation infrastructure, road infrastructure, irrigation infrastructure, drinking water infrastructure, wastewater infrastructure, telecommunication infrastructure, electricity infrastructure, and oil and gas infrastructure.

Infrastructure Indicators

Infrastructure indicators are metrics that assess performance outcomes, elucidate accomplishments, are quantifiable, and are pertinent to the development efforts undertaken by local governments (Handayani & Sophianingrum, 2019)

K-Means Method

The K-Means clustering technique is a non-hierarchical strategy that seeks to split data into two or more groups by utilizing the mean value as the centroid of each cluster (Oktarina et al., 2020). The K-Means algorithm is the most prevalent and extensively utilized clustering technique. It aims for a well delineated split that reduces the distance between clusters (Wang & Gao, 2019). The centroid of the cluster may be determined using equation (1), specifically:

$$J(x, v) = \sum_{k=1}^{n_i} \sum_{i=1}^{c} \mu_{ik} d^2_{ik} \quad \text{.....(1)}$$

where *(ni)* is the number of objects belonging to the *(i)* group, c is the number of clusters and µik is the distance value between objects to the i-th cluster centre.

Average Linkage Method

(Widyadhana et al., 2021) The Average Linkage approach involves merging two clusters with the minimal average distance until all items are consolidated into a single giant cluster. The mean distance between each pair of objects from two clusters is minimized following their amalgamation into a new cluster based on the intergroup relationships. Calculation of clusters utilizing the Average Linkage technique as delineated in equation (2):

$$d_{(AB)C} = \frac{\sum_u \sum_v d_{uv}}{N_{(AB)} N_C} \quad \text{.....(2)}$$

where duv is the distance between objects in cluster AB and object v in cluster C, N(AB) is the number of objects in cluster AB and NC is the number of objects in cluster C.

III. METHODS

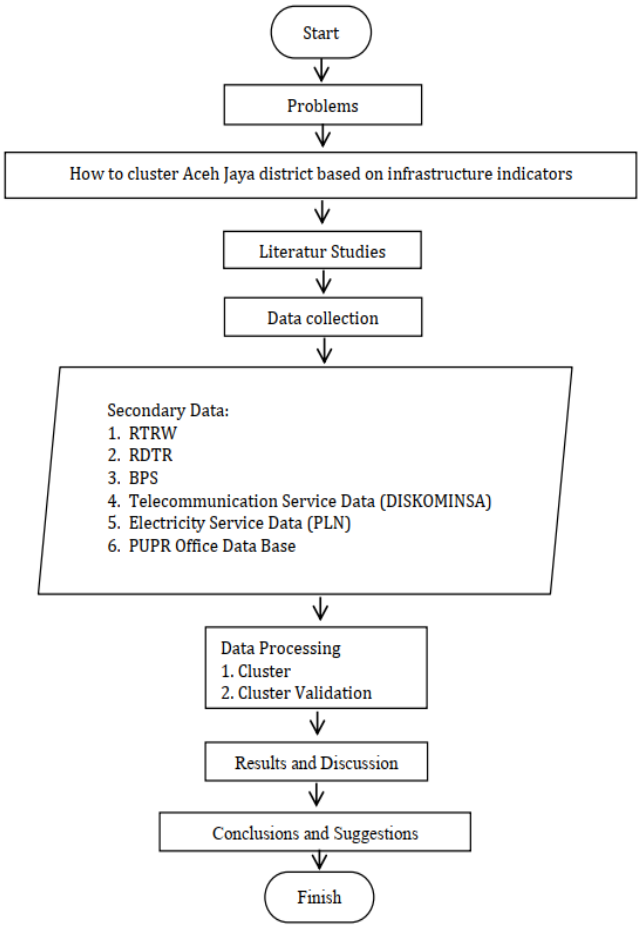

Fig 1. Research flow chart

Research Location

This research is located in Aceh Jaya Regency which is located on the west coast of Aceh. Administratively, Aceh Jaya Regency has regional boundaries, namely the North bordering Aceh Besar Regency and Pidie Regency, the South with West Aceh Regency and the Indian Ocean, the West with the Indian Ocean and the East bordering West Aceh Regency. Aceh Jaya Regency itself consists of 9 sub-districts and 172 Gampong. The nine sub-districts are Teunom, Pasie Raya, Panga, Krueng Sabee, Setia Bakti, Sampoiniet, Darul Hikmah, Jaya, Indra Jaya (Aceh Jaya Regency Spatial Plan 2014-2034). For more details on the research location, please see Figure 2.

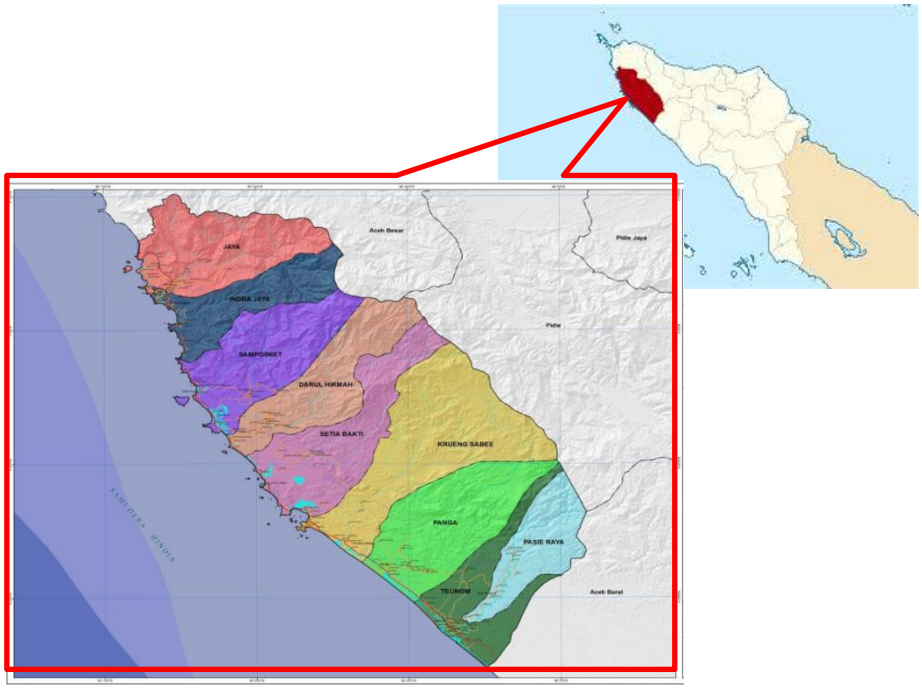

Fig 2. Research Location

Research Variables

This analysis utilizes infrastructure variables derived from the indicators in the Rencana Pembangunan Kabupaten Aceh Jaya 2023-2026. This research presents the selected indicators based on the infrastructure scope outlined in Presidential Regulation No. 67 of 2005 of the Republic of Indonesia, as seen in the following table:

Table 1. Research Variables

Infrastructure Variables	Indicators
X1	Irrigation Area (Ha)
X2	House with Access to Drinking Water
X3	Sanitised Households
X4	Drainase (Km)
X5	District Road Length (Km)
X6	Livable House (Unit)
X7	Telecommunication Served
X8	Electricity Served

Source: Research Results

Data Analysis

In this study the data used is secondary data, namely data obtained from related agencies, namely from Dinas Pekerjaan Umum dan Perumahan Rakyat Kabupaten Aceh Jaya, Perusahaan Umum Daerah (PERUMDA) Tirta Mon Mata Aceh Jaya Regency, Dinas Komunikasi dan Persandian (DISKOMINSA) Aceh Jaya Regency, PLN Aceh Jaya Regency. The analysis method in this study uses clustering analysis to obtain groups of areas with similar characteristics comparing them with two clustering methods namely K-Means and Average Linked. Data processing was done with the help of Statistical Product and Service Solution (SPSS) 27 software.

IV. RESULTS AND DISCUSSION

Cluster analysis assumptions (Representative Sample)

The cluster analysis assumption test using the Kaiser-Mayer-Olkin (KMO) test was carried out twice. The first test results with eight variables resulted in a KMO value below 0.5 so that the variable was eliminated by looking at the smallest MSA value, namely the X1 variable. After being tested again with seven variables,

the KMO value was obtained above 0.5. The results of the KMO assumption retest can be seen in the following results:

Table 3. KMO Test Results

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0,611
Bartlett's Test of Sphericity	Approx. Chi-Square	71,011
	df	21
	Sig.	0,000

Source: Research Results

The results of retesting with variables X2, X3, X4, X5, X6, X7, and X8 obtained a KMO value of 0.611, which means that the data used can be a representative sample.

Data standardisation

Data units with significant differences must be considered in a study. Data that has significantly different units must be standardised by converting the previous data to Zscore. If the zscore value with the amount of data below 80 has a value ≥ 2.5, it means that the data used is outliar so the data must be standardised. The results of the Zscore value test with the number of samples below 80 do not occur outliar because the resulting value is not ≥ 2.5, so there is no need to standardise the data.

Clustering with K-Means Method

Cluster analysis with the K-Means method in this study determined the number of clusters to be formed as many as 3 clusters, the clusters can be seen in table 2 below:

Table 2. Results of cluster analysis with the K-Means method

CLASTER	DISTRICTS
1	Teunom
	Krueng Sabee
2	Pasie Raya
	Panga
	Setia Bakti
	Sampoiniet
	Darul Hikmah
	Indra Jaya
3	Jaya

Source: Research Results

From the cluster results, it is found that the highest cluster centre results are in clusters 1 and 3, which means that the level of infrastructure development served by drinking water, sanitated households, drainage length, road length, livable houses, electricity services is in cluster 1, namely Teunom and Krueng Sabee sub-districts, while the level of infrastructure development served by telecommunications is in cluster 3, namely Jaya sub-district. The final cluster centre results can be seen in the following table :

Table 3. Final Cluster Centers

Final Cluster Centers			
	Cluster		
	1	2	3
Drinking Water Served	3120	855	1207
Sanitised Households	4436	2017	4211
Drainage Length	141,10	103,11	110,19
Road Length	128,87	57,47	79,45
Livable House	4637	2220	4423
Telecommunication Served	7541,518	5556,861	13102,000
Electricity Services	6624	2789	4916

Clustering with the Average Linkage Method

Clustering analysis using the Average Linkage method can be seen from the dendrogram results. From the dendrogram results, the clusters formed are 2 clusters, namely cluster 1 consisting of Darul Hikmah, Indra Jaya, Pasie Raya, Setia Bakti, Sampoiniet and Panga sub-districts. While cluster 2 consists of Teunom, Krueng Sabee and Jaya sub-districts.

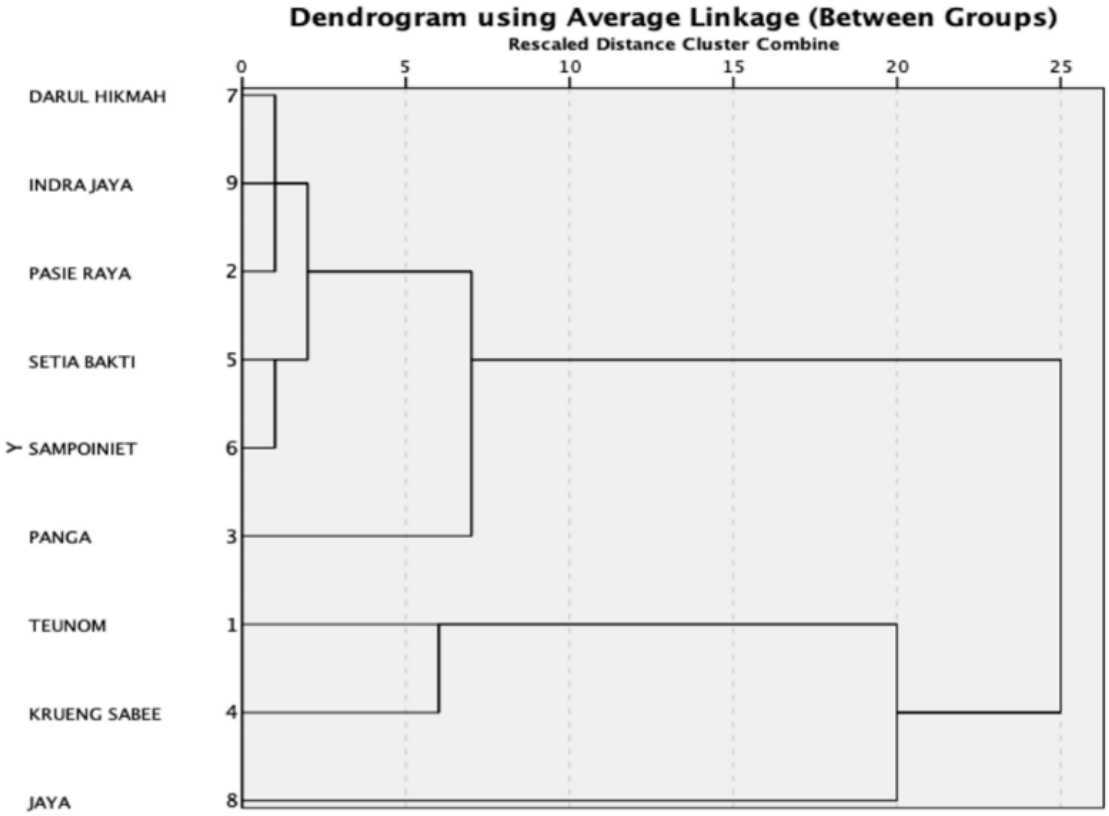

Fig 3. Dendogram Average Lingkage

Cluster members with Average Linkage metrode are as follows:

Table 4. Results of cluster analysis with K-Means method

CLASTER	DISTRICTS
1	Teunom
	Krueng Sabee
	Jaya
2	Panga
	Setia Bakti
	Sampoiniet
	Darul Hikmah
	Indra Jaya
	Pasie Raya

Source: Research Results

Cluster Validation and Profiling

From the results of cluster validation using the Silhouette Coefficient formula, it is obtained that the best cluster is the cluster with the Average Linkage method because the resulting Silhouette Coefficient value is closest to 1. The results of the Silhouette Coefficient calculation are as follows:

Table 5. Score Silhouette Coefficient

Klaster	Silhouette Coefficient
K-Means	0.636
Average Linkage	0.736

Source: Research Results

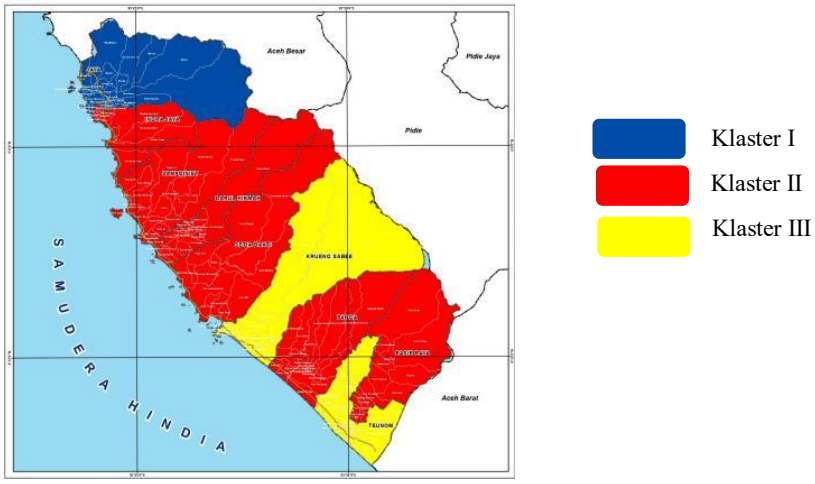

Fig 4. K-Means Method Cluster Visualisation

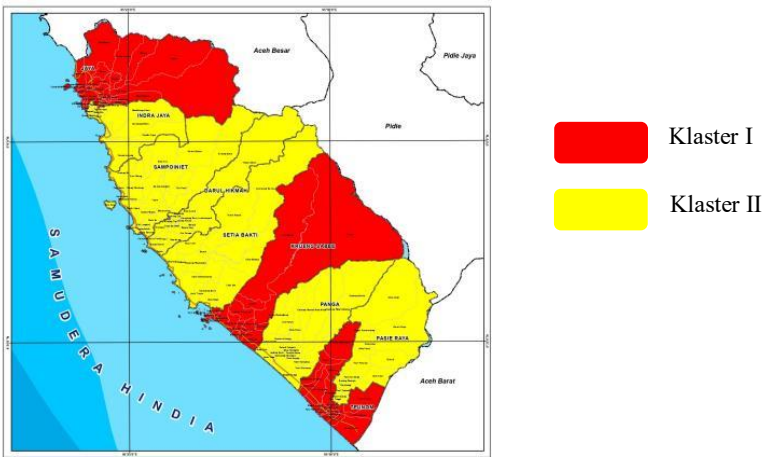

Fig 5. Cluster Visualisation of Average Linkage Method

V. CONCLUSION

As a result of the above research, it is found that clustering with the K-Means method can be divided into 3 clusters, namely the first cluster consists of Jaya District, the second cluster consists of Krueng Sabee and Teunom Districts while the third cluster consists of Indra Jaya, Sampoiniet, Darul Hikmah, Setia Bakti, Panga and Pasie Raya Districts. The clusters that are formed are spread out because the nature of the cluster formed is in accordance with the level of homogeneity of the object being assessed, namely the level of infrastructure development in these sub-districts. Likewise, the clusters formed by the Average Linkage method, namely 2 clusters, are also spread as in the first cluster consisting of Teunom, Krueng Sabee, and Jaya sub-districts while the second cluster consists of Panga, Setia Bakti, Sampoiniet, Darul Hikmah, Indra Jaya, and Pasie Raya sub-districts.

In the results of the study it was also found that the inequality of infrastructure distribution in Aceh Jaya District can be seen from the results of the level of development of electricity infrastructure, livable houses, road length, drainage, sanitised households, served drinking water and telecommunications in cluster 1 is higher than cluster 2 so that what is needed is equal distribution of infrastructure and it is necessary to determine strategies and priorities for sub-districts in the second cluster, namely Panga District, Setia Bakti, Sampoiniet, Darul Hikmah, Indra Jaya and Pasie Raya sub-districts where infrastructure development indicators are inadequate considering that these sub-districts also require adequate infrastructure for the continuity of economic cycles, community welfare and disaster response areas.

REFERENCES

[1] Amelia,S.,Maulani,A.,(2022)Klasterisasi Kota Berdasarkan Indikator Indikator Pembangunan Manusia Dengan Metode Avarange Linkage. ***Jurnal Unirow***,Vol 04, No.02,69-74.p-ISSN 2656-6303–e-ISSN 2656-9876.

[2] Ardinugroho, N. S., Handayani, W., Sophianingrum M. (2019). Alternatif Indikator Infrastruktur Di Kota Semarang: Identifikasi Menuju Pengukuran Berbasis Kinerja. ***Jurnal RIPTEK***,Vol. 13 (2) 137-146,

[3] Bahri, M. I., Kismawadi, E. R., & Agustina, T. S. (2024). Between Sympathy and Hostility: Acehnese Attitudes towards Rohingya. ***Analisa: Journal of Social Science and Religion***, *9*(2).

[4] Bahri, M. I., & Ridha, A. E. (2024). *Manajemen Organisasi: Sektor Publik*. Gue Media Group.

[5] Bahri, M. I., Sirojuzilam, S. S., Rujiman, R. R., & Kismawadi, E. R. (2021). The Influence of Youth in Regional Development Through the Management of Aceh's Special Autonomy Fund in Langsa City. ***E-Mabis: Jurnal Ekonomi Manajemen Dan Bisnis***, *22*(1). https://doi.org/10.29103/e-mabis.v22i1.650

[6] Badruddin, S., Achmad, A., Fadhly, N., (2022) Model Klasterisasi Wilayah Kota Bedasarkan Indikator Infrastruktur Untuk Perencanaan Struktur Ruang Berkelanjutan. ***JRSP :Jurnal Arsip Rekayasa Sipil dan Perencanaan,*** 5(3), 222-235, https://doi.org/10.24815/jarsp.v5i3.26749

[7] Bu., J., Liu, W., Pan, Z., & Liang, K., (2020) Comparative Study of Hydrochemical Classification Based on Different Hierarchical Cluster Analysis Methods, ***International Journal Enviorment Research and Public Health, China***, https://www.mdpi.com/1660-4601/17/24/9515/pdf

[8] Cahyani, I. R., & Bahri, M. I. (2022). Change Management and Special Autonomy Paradigm: Opportunities and Challenges of Human Resources in Digital Literacy Implementation in Indonesian Regional Libraries. ***Congress of Southeast Asian Librarians (CONSAL) XVIII.***

[9] Dinda Puspitaa, Ellida Novita Lidya, Firdasari, & M Irvanni Bahri. (2022). Analysis Of Successful Factors In The Halal Tourism Facilities Development Project. ***Ihtiyath : Jurnal Manajemen Keuangan Syariah***, *6*(2).

[10] Fathia, A. N., Rahmawatia, R., & Tarno., (2016) Analisis Klaster Kecamatan Di Kabupaten Semarang Bedasarkan Potensi Desa Menggunakan Metode Ward dan Single Linkage. Jurnal Gaussian, Vol 5, Nomor 4, 801-810

[11] Handayani, N. S., & Sophianingrum, W. (2019). Alternatif Indikator Infrastruktur Di Kota Semarang: Identi-Fikasi Menuju Pengukuran Berbasis Kinerja. In ***Jurnal Riptek*** (Vol. 13, Issue 2).

[12] Isya, M., Fadhly, N., & Saputra, A. (2021). Multi Criteria Analysis Of Road Handling Type At Banda Aceh - Meulaboh National Road Segment Indonesia. ***Journal Of Physics: Conference Series***, *1845*(1).

[13] Jiawei Han, M. K. And J. P. (2012). *Data Mining: Concepts And Techniques, Third Edition - Books24x7*. Morgan Kaufmann Publishers.

[14] Kabupaten Aceh Jaya, Rencana Pembangunan Kabupaten (RPK) Aceh Jaya Tahun 2023-2026. BAPERRIDA Kabupaten Aceh Jaya, 2023

[15] Kabupaten Aceh Jaya, BAPPERIDA. 2023. *Rencana Pembangunan Kabupaten (RPK) Aceh Jaya Tahun 2023-2026*: Kabupaten Aceh Jaya

[16] Kabupaten Aceh Jaya, DISKOMINSA. 2023. *Kondisi Telekominukasi Kabupaten Aceh Jaya*:Kabupaten Aceh Jaya

[17] Kabupaten Aceh Jaya, PUPR. *Qanun Kabupaten Aceh Jaya Nomor 9 Tahun 2014 Tentang Rencana Tata Ruang Wilayah Kabupaten Aceh Jaya Tahun 2014 2034*: Kabupaten Aceh Jaya

[18] Kabupaten Aceh Jaya,PUPR.2023.*Database Dinas Pekerjaan Umum dan Perumahan Rakyat*

[19] Masdi, & Iskandar, I. (2021). Causality Relationship Between Infrastructure, Investment And Grdp Per Capita In Aceh. ***Jurnal Ekonomi Dan Pembangunan***, *12*(2).

[20] Oktarina, C., Notodiputro, K. A., & Indahwati, I. (2020). Comparison Of K-Means Clustering Method And K-Medoids On Twitter Data. ***Indonesian Journal Of Statistics And Its Applications***, *4*(1).

[21] Sanjaya, I. P. A. (2018). Perkembangan Infrastruktur Di Indonesia. *17 Oktober 2018*.

[22] Tambunan, A. A., Purwoko, A., Rambe, W., & Bahri, M. I. (2024). *Strategi Pemberdayaan Masyarakat Melalui Pengembangan Desa Wisata Kampoeng Lama* (Vol. 08, Issue 01).

[23] Undang - Undang Nomor 26 Tahun 2007 tentang Penataan Ruang

[24] Wang, T., & Gao, J. (2019). An Improved K-Means Algorithm Based On Kurtosis Test. ***Journal Of Physics: Conference Series***, *1267*(1). Https://Doi.Org/10.1088/1742-6596/1267/1/012027

[25] Widarjono, A. (2010) Analisis Statistika Multivariat Terapan. Edisi Pertama. Yogyakarta : UPPSTIM YKPN.

[26] Wu, X., Gan, L., Yuan, S.. Rui, D. (2022) A preliminary study on wellbore flow interpretation of fiber optic vibration signals based on K-means clustering algorithm. SN Appl. Sci. 4, 233.

[27] Widyadhana, D., Hastuti, R. B., Kharisudin, I., & Fauzi, F. (2021). Perbandingan Analisis Klaster K-Means Dan Average Linkage Untuk Pengklasteran Kemiskinan Di Provinsi Jawa Tengah. *Prisma, Prosiding Seminar Nasional Matematika*, *4*, 584–594.