Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.4, August 2025, pp. 1544~1550
e-ISSN: 2548-6861

Hierarchical Clustering of Education Indicators in Papua Island: A Ward's Method Approach

Ifa Prastika [1*], Devni Prima Sari [2*]
*Department of Mathematics, Universitas Negeri Padang
[2] devniprimasari@fmipa.unp.ac.id

Article Info

Article history:

Received 2025-06-03
Revised 2025-07-17
Accepted 2025-07-30

Keyword:

Clustering,
Education,
Hierarchical,
Ward.

ABSTRACT

Education development aims to ensure inclusive, equitable education and increase learning opportunities for all Indonesian citizens. Papua Island is still not an island with a high education level; data on education indicators indicate this in each Regency / City on the island of Papua, with a value below the national average. Identifying districts/cities is needed to improve education, so clustering is carried out using the *Ward* method. This research aims to group and map regencies/cities on the island of Papua based on education indicators. The results of this study are expected to be a consideration and benchmark for the government in making decisions regarding education in districts/cities on the island of Papua, considering the region's characteristics. This is an applied research with the data type used, namely secondary data on education indicators in Papua Island in 2022. Data sources are obtained from the official website of the Central Bureau of Statistics of each province on the island of Papua. Four education indicators are taken into account in this research, namely the School Participation Rate (SPR), the Gross Enrollment Rate (GER), the Net Enrollment Ratio (NER), and the Average Years of Schooling (AYS), which are then detailed into 10 variables. *The cluster* analysis process uses *Euclidean* distance and *cluster* validation using the *Dunn Index*. The results showed that 3 *clusters* formed. *Cluster* 1 consists of 27 districts/cities; this first group is classified as a high level of education. Cluster 2 consists of 7 districts/cities with a medium level of education, and Cluster 3 has eight districts/cities with a low level of education—cluster results based on the highest *Dunn Index* validation value of 0.414.

I. Introduction

Education can improve the quality of human resources better and modernly, if the education system runs properly [1]. The goal of education development is, of course, to ensure inclusive, equitable education and increase the desire to learn for all Indonesian citizens [2] . Papua Province is an area that mostly consists of wilderness and mountains. Several education problems arise, ranging from difficulties in recruiting teaching staff, poor security factors, and geographical conditions that result in difficult access to education for residents [3].

Based on data from the Central Bureau of Statistics (2024), the province with the lowest School Participation Rate (SPR) on Papua Island is Central Papua Province with a SPR in the 7-12 year age group of 74.63 percent, a school participation rate (SPR) in the 13-15 year age group of 68.42 percent, and a SPR in the 16-18 year age group of 47.65 percent, which is still below the average achievement of national standards. Therefore, it is essential to analyse the clustering and mapping of districts/cities on Papua Island based on education indicators to determine which districts/cities need improvement in education problems.

Cluster analysis is defined as grouping objects or cases based on the similarity of their characteristics. Objects or cases in each cluster tend to have similarities and differ significantly from objects in different clusters [4]. Two procedures can be used for cluster analysis, namely hierarchical and non-hierarchical clustering. In hierarchical clustering, the process begins with the assumption that each object is a cluster. Two or more objects are combined based on high homogeneity measured by the closest distance; This process continues until the cluster forms a clear level tree (dendrogram) between objects, from the most similar to the least similar [5]. In contrast to hierarchical clustering, non-hierarchical clustering starts by specifying the desired number of clusters. Hierarchical clustering has several methods, including the Single Linkage, Complete Linkage, Centroid, Average Linkage, and Ward methods [6].

Ward's method is part of the variance method that has a vision to obtain cluster results by minimizing the variance within the cluster. This method belongs to a hierarchical clustering procedure that is often preferred over other methods because the Ward method provides stable results and consistent dendrogram interpretation capabilities [7]. The distance formed between two clusters is called the Sum of Squared Errors (SSE). In this study, the distance measure used is the Euclidean distance. The Dunn Index calculation is used to determine the best number of clusters using the Ward method [8].

Some studies that have applied the ward method include research by Dewi et al, who conducted a cluster analysis for grouping regencies/cities in Maluku Province based on education indicators using the Ward method [9]. Using the same method and different problems, research conducted by Rahmadani and Salma used the Average Linkage and Ward methods in clustering the welfare of West Sumatra in 2021 [10].

This study focuses on clustering the districts and cities in Papua Island based on key educational indicators: School Participation Rate (SPR), Gross Enrollment Rate (GER), Net Enrollment Ratio (NER), and Average Years of Schooling (AYS). These indicators are further detailed into ten standardised variables and analyzed using hierarchical clustering with Ward's method. The findings are expected to inform education policy by identifying clusters of regions with similar needs and development priorities.

II. METHODS

This study is categorised as applied research, aiming to implement clustering theory in a real-world context, specifically in analysing educational disparities across provinces on Papua Island. of the several hierarchical methods, the ward method is often chosen because it provides stable results and consistent interpretation of the resulting dendrogram[7]. In addition, in the ward method, each step in the cluster merging (agglomeration) process is always based on the minimum distance between clusters[11]. The objective is to group provinces based on similarities in their education indicators to support more targeted educational policies. The data used in this research were obtained from the official website of the Central Bureau of Statistics (Badan Pusat Statistik/BPS) for all provinces located on Papua Island, covering the most recent data available for 2022. The research methodology consists of the following steps, as illustrated in Figure 1.

A. *Euclidean Distance*

Euclidean Distance is one of the most commonly used measures to calculate the similarity or dissimilarity between two data points in a multidimensional space. It represents the straight-line distance between two objects in a p-dimensional space and is widely applied in clustering, classification, and pattern recognition.

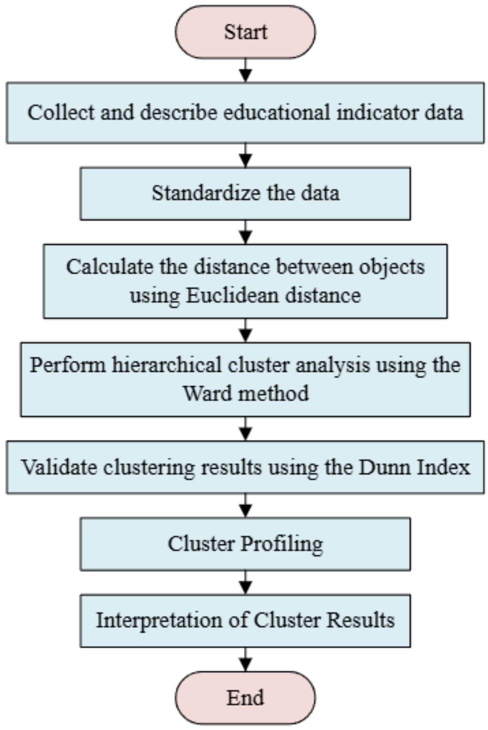

Figure 1. Research methodology

Mathematically, the Euclidean distance between two objects i and j, each with p attributes, is defined as follows [12]:

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

Where:

- d_{ij} is the distance between object i and object j,
- x_{ik} and x_{jk} are the values of the k-th variable for objects i and j, respectively,
- p is the number of variables (dimensions).

This method is sensitive to differences in scale; therefore, standardization of data is typically required before applying Euclidean distance in cluster analysis.

B. Ward's Method

Ward's method, introduced by Joe H. Ward Jr. in 1963, is an agglomerative hierarchical clustering technique that minimizes the within-cluster variance. Ward's method was selected in this study due to its primary advantage in minimizing the total variance within clusters, or more precisely, reducing the Sum of Squared Errors (SSE) at each stage of the hierarchical merging process. Ward's method merges the pair of clusters that results in the smallest increase in SSE, thereby producing clusters that are internally homogeneous and clearly separated from one another[15]. In contrast to other linkage methods, Single linkage relies on the minimum distance between points and tends to produce the so-called chaining effect, where distant objects are continuously merged into a single elongated cluster, resulting in a structure that lacks compactness [16]. Complete linkage, on the other hand, uses the maximum distance between points, which often leads to more compact clusters but is highly sensitive to outliers, potentially splitting natural clusters inappropriately [17]. Average linkage, though a compromise between the two, remains less effective in controlling internal variation, especially when dealing with heterogeneous or multivariate data [18]. At each step, it merges the pair of clusters, which results in the smallest increase in the sum of squared errors (SSE), which measures the total deviation of data points from their cluster mean. The SSE is calculated using the following formula [13]:

$$I_{AB} = \frac{n_A n_B}{n_A + n_B}\left((\bar{x}_A - \bar{x}_B)'(\bar{x}_A - \bar{x}_B)\right)$$

Where SSE_{ij} is increase in error when merging cluster i and j, y_i is the mean vector of cluster i, and y_j is mean vector of cluster j. Ward's method ensures the formation of compact and homogeneous clusters.

III. RESULTS AND DISCUSSION

A. Descriptive Analysis

This study used four key education indicators to evaluate the educational conditions across districts in Papua Island. These indicators were broken down into ten standardized variables (Z1–Z10) to capture different school participation, enrollment, and educational attainment aspects. The details of each indicator and its corresponding variables are presented in Table I.

In this study, 10 education variables are used that reflect the dimensions of access, timeliness, and long-term educational attainment at the district/city level in Papua Island. The first to third variables are the School Participation Rate (SPR) for age groups 7-12 years, 13-15 years, and 16-18 years. The SPR for ages 7-12 is used to measure children's access to primary school, which is critical in detecting potential early dropouts. SPR age 13-15 reflects the attainment of junior secondary School, which is an important indicator of the success of the 9-year compulsory education program. Meanwhile, the SPR aged 16-18 reflects the level of continuation of education to the upper secondary level , which determines students' readiness for the world of work or higher education.

TABLE I
DESCRIPTIVE STATISTICS OF EDUCATIONAL INDICATORS

Main Indicator	Variable Code	Description
School Participation Rate (SPR)	Z1	SPR for ages 7–12 years
	Z2	SPR for ages 13–15 years
	Z3	SPR for ages 16–18 years
Gross Enrollment Rate (GER)	Z4	GER for Primary School (SD) or equivalent
	Z5	GER for Junior Secondary School
	Z6	GER for Senior Secondary School
Net Enrollment Ratio (NER)	Z7	NER for Primary School
	Z8	NER for Junior Secondary School
	Z9	NER for Senior Secondary School
Average Years of Schooling (AYS)	Z10	Average Years of Schooling (AYS)

The fourth through sixth variables are the Gross Enrollment Rate (GER) for primary, junior secondary, and senior secondary levels. GER measures the capacity of the education system without limiting the age of students, so a high GER value (>100%) can reflect the success of educational inclusion but also indicate a delay in school entry or grade repetition. The primary GER reflects a region's ability to accommodate primary school-age children, while the junior secondary GER and senior secondary GER assess the inclusiveness of education at the junior secondary and senior secondary levels. These three GER are important to see if the school capacity at each level is sufficient or if there are still gaps.

Next, the seventh to ninth variables are the Net Enrollment Ratio (NER) for elementary, junior high, and high school. Unlike SPR, GER, only counts students who attend school according to their ideal age. Primary NER shows how many students enter primary school on time, junior high school NER measures the efficiency of the system in retaining students on time in junior high school, and senior high school NER reflects the successful transition and graduation of students without delay in senior high school. Age accuracy is very important to reflect the efficiency and sustainability of quality education.

Finally, the tenth variable is the Average Years of Schooling (AYS) which shows the average number of years that the population aged 25 years and above has received formal education. AYS is an accumulative indicator that shows the long-term results of the education system in a region, as well as reflecting the quality of human resources. This variable is very important in evaluating educational achievements across generations and regional economic potential. Together, these ten variables provide a comprehensive picture of the condition of education in Papua, covering aspects of access, fidelity and outcomes, making them highly relevant for use in cluster analysis of education inequality

The results of the descriptive analysis for these variables are presented in Table II. Table II presents the descriptive statistics of the educational indicators (Z1–Z10) across 42 districts/cities in Papua Island. The data show variations in mean, minimum, and maximum values, reflecting differences in educational conditions across regions. This variation supports cluster analysis to group areas with similar academic profiles.

B. Data Standardization

Education indicator data has different units, so a data standardization process is necessary. The data is standardized using the Z-score calculation. Z-score is mathematically formulated [12] :

$$Z = \frac{X - mean(X)}{SD\ (X)}$$

Description:

Z : Data standardization result

X : Data used

$mean(X)$: Average of the data

$SD\ (X)$: Standard deviation of the data

TABLE II
DESCRIPTIVE STATISTICS OF EDUCATIONAL INDICATORS

Variables	Min	Max	Average	Standard Deviation
Z1	51,61	99,95	85,89	12,89
Z2	31,98	98,55	85,22	15,92
Z3	27,71	97,12	67,76	17,90
Z4	57,06	125,39	100,19	16,81
Z5	37,06	107,78	81,93	18,72
Z6	15,12	143,48	78,26	29,71
Z7	51,96	98,11	84,78	12,97
Z8	16,05	86,75	60,19	17,39
Z9	9,99	71,39	47,51	18,93
Z10	1,71	11,84	6,81	2,68

C. Calculating Euclidean Distance

In this stage, the Euclidean distance was calculated between all pairs of districts/cities in Papua Island using the standardized educational indicators (Z1–Z10). Each district/city is treated as an observation in a 10-dimensional space, and the distance between each pair reflects the level of similarity or dissimilarity in their educational profiles.

Suppose we want to calculate the *Euclidean* distance between Merauke Regency and Jaya Pura Regency (object 1 and object 2). Then:

$$d_{1,2} = \sqrt{\sum\nolimits_{k=1}^{p} (z_{ik} - z_{jk})^2}$$

$$= \sqrt{(Z_{1,1} - Z_{2,1})^2 + (Z_{1,2} - Z_{2,2})^2 + \cdots + (Z_{10,1} - Z_{10,2})^2}$$

$$= \sqrt{(0{,}61 - 0{,}14)^2 + (0{,}75 - 0{,}24)^2 + \cdots + (0{,}91 - (-0{,}31))^2}$$

$$= \sqrt{(0{,}61 - 0{,}14)^2 + (0{,}75 - 0{,}24)^2 + \cdots + (0{,}91 - (-0{,}31))^2}$$

$$d_{1,2} = 3{,}35$$

The distance between the data in Merauke and Jaya Pura districts is 3.35. The result is a 42 × 42 distance matrix, where each cell $d_{i,j}$ represents the distance between district *i* and district *j*. This matrix forms the foundation for the subsequent hierarchical clustering process.

D. Clustering Using Ward's Method

This study applies Ward's hierarchical agglomerative clustering technique that minimizes the increase in the SSE at each merging step. All 42 districts/cities in Papua Island are initially treated as individual clusters. The two clusters with the smallest increase in SSE are combined in each iteration.

The clustering was performed using RStudio. After computing the Euclidean distance matrix, the first merge occurred between Nabire and Biak Numfor (Clusters 4 and 6), with the smallest distance of 0.54, forming a new cluster and reducing the total to 41. SSE for merging two clusters A and B is calculated as:

$$SSE_{4,6} = \frac{n_4 n_6}{n_4 + n_6} [(\bar{z}_4 - \bar{z}_6)'(\bar{z}_4 - \bar{z}_6)]$$

In this case:

$$SSE_{4,6} = \frac{(1)(1)}{(1) + (1)} (0.54) = 0.27$$

In the next step, Fakfak (Cluster 31) is merged with Manokwari (Cluster 35), with the lowest SSE value of 0.27. The matrix is then updated and continues until all objects are merged. This agglomerative process ensures that the most similar clusters are combined at each stage.

The complete clustering process is visualized through a dendrogram (Figure 2).

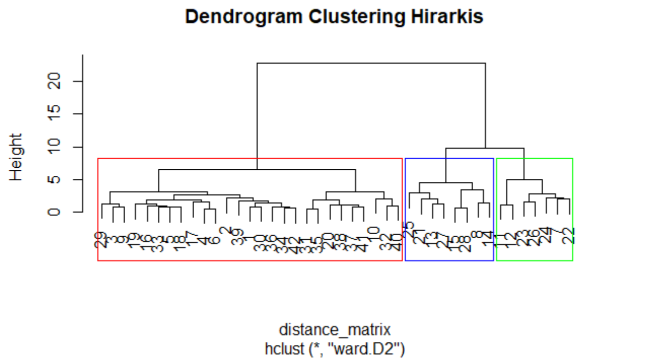

Figure 2. Dendrogram of the grouping of regencies/cities in Papua Island in 2022 based on education indicators.

E. *Validation Clustering result using the Dunn Index*

Cluster validity indices often yield different recommendations for the optimal number of clusters, depending on the characteristics of the data and the emphasis of the index. For instance, Dunn favors well-separated clusters, while Silhouette combines compactness and separation[14]. The higher the Dunn Index value, the better the clustering [8]. The number of clusters that maximizes the Dunn Index is taken as the optimal number of clusters. The results of the Dunn Index calculation for grouping districts /cities in Papua Island based on education indicators are presented in Table III.

TABLE III
DUNN INDEX CALCULATION RESULTS

Number of Clusters	Dunn Index
2	0,255
3	**0,414**
4	0,223
5	0,223
6	0,299

Table 3 shows that the highest value of the Dunn Index calculation is at 3 clusters. This means the best number of clusters in this Ward method grouping is 3.

A good clustering is one for which the inter-cluster distances are large and the intra-cluster distances are small. The ratio of minimum inter-cluster distance to maximum intra-cluster diameter can be used to evaluate such clustering structures[15]. Ward method was used due to its ability to minimize variance within clusters. The clustering was evaluated using internal validity indices, and the best number of clusters was selected based on interpretability and consistency with regional characteristics.

F. *Cluster Profiling*

The results of *clustering* districts/cities in Papua Island based on education indicators using the *ward's* method can be seen in the dendrogram (Figure 2). Because the optimal number of clusters is 3, each *cluster is* cut with several lines. The red line represents *cluster* 1, the green line represents *cluster* 2, and the blue line represents *cluster* 3. The detailed members of each *cluster* are presented in Table IV.

TABLE IV
CLUSTER MEMBERSHIP

Cluster	Number of Members	Member
Cluster 1	27	Merauke, Jayawijaya, Jaya Pura, Nabire, Yapen Islands, Biak Nufor, Mimika, Boven Digoel, Sarmi, Keerom, Waropen, Supiori, Mamberamo Raya. Kota Jaya Pura, Fakfak, Kaiman, Teluk Wondama, Teluk Bintuni, Manokwari, South Sorong, Sorong, Raja Ampat, Tambrauw, Maybrat, South Manokwari, Arfak Mountains, Sorong City.
Cluster 2	7	Mappi, Asmat, Memberamo Tengah, Dogiyai, Yalimo, Lani Jaya, Panilai districts.
Cluster 3	8	Nduga, Puncak, Intan Jaya, Yahukimo, Tolikura, Deiyai, Puncak Jaya, Bintang Mountains.

G. *Interpretation of Cluster Result*

Cluster interpretation is based on the average value of variables in each *cluster*. The results of the average variables in each *cluster* are presented in Table V.

TABLE V
AVERAGE OF CLUSTER VARIABLES

Variables	Cluster1	Cluster2	Cluster3
Z1	0,61	-0,32	-1,79
Z2	0,62	-0,45	-1,69
Z3	0,63	-0,59	-1,61
Z4	0,57	-0,48	-1,50
Z5	0,48	-0,008	-1,61
Z6	0,58	-0,80	-1,27
Z7	0,62	-0,38	-1,76
Z8	0,51	-0,24	-1,53
Z9	0,59	-0,69	-1,40
Z10	0,59	-0,78	-1,32

Table V shows that the average value of the variables of *each* cluster is used to identify the features of *the* three clusters generated.

1) Cluster 1 Interpretation: Cluster 1 is characterized by the highest 16-18 years SPR (Z3) - indicating that most teenagers are still attending high school. The senior secondary school GER (Z6) also peaks, indicating that there is relatively sufficient senior secondary school capacity in the region. In the age-appropriateness domain, the primary school enrollment rate (Z7) is quite high, indicating that most children enter primary school on time. The Average Years of Schooling metric (Z10) is above the overall average, corroborating evidence of good educational accumulation. The main weakness is the junior secondary school enrollment rate (Z5) - the lowest among the cluster variables - indicating that there are still students beyond the 13-15 age range who attend junior secondary school due to grade repetition or delay. Implications: Accelerated junior high school graduation and age adjustment programs are the focus, while interventions at the senior high school level are geared towards quality improvement.

2) *Interpretation of Cluster 2:* The dominant variable in cluster 2 is the NER for Junior Secondary School (Z8), which indicates that most adolescents 13 to 15 years old are attending school at the appropriate level. In contrast, the GER for Junior Secondary School (Z5) is the lowest, which means that the capacity of junior secondary schools has not reached all students outside the ideal age range (e.g. older students who have dropped out of school). The SPR 16 to 18 and GER for Senior Secondary School values are at a moderate level, indicating that the transition to upper levels has not been fully optimized. Implication: Policy priority should be on expanding the capacity of junior secondary schools (classrooms, teachers), as well as equalization programs for late entrants.

3) *Cluster 3 shows* the GER for Senior Secondary School (Z6) but the lowest SPR for ages 7-12 years (Z1). This means that many people are still in senior high school despite being of inappropriate age, while the participation of primary school-age children is very low. This combination indicates unequal access to basic education and chronic school delays. APM values at all levels are also below average, confirming the widespread problem of age inappropriateness. Implications: urgent interventions in basic education to provide schools near remote settlements, transportation scholarships, and illiteracy reduction programs as well as addressing over-age students through accelerated classes or package education.

Distinctive Features of Each Cluster

Cluster 1 - High (Education Access and Outcomes Good)

Distinctive Features:

High school enrollment rates for 16-18 year olds (Z3), indicating that a large proportion of the teenage population is attending school and continuing on to higher education.The gross enrollment rate (GER) of senior high school is high (Z6), indicating that access to senior high school education is relatively good in the region. The primary school enrollment rate (Z7) is also high, which means that many children aged 7-12 years are attending school at the appropriate level. Average Years of Schooling (Z10) indicates that residents in this cluster generally have higher levels of education, reflecting a more advanced level of development in the education sector. Examples of Regions in This Cluster:

Jayapura: As the capital of the province, it has better education facilities, more adequate infrastructure, and higher educational attainment than other regions in Papua.

Sorong City: A developing city with better access to education and school facilities.

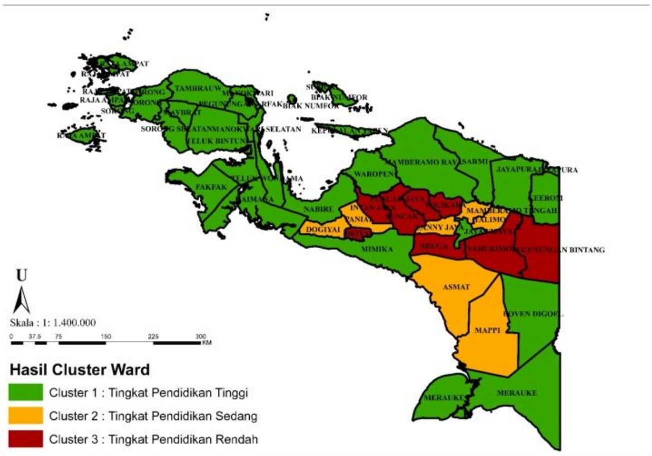

Figure 3. Map of regencies/cities in Papua Island in 2022 based on education indicators.

Cluster 2 - Medium (Adequate Education with Certain Challenges).

Distinctive Features:

NER for Junior Secondary School is high (Z8), indicating that a large proportion of the population aged 13-15 years attend junior secondary school at the appropriate level.

The junior secondary GER is lower (Z5), indicating that some junior secondary students are delayed in entering school. Senior high school enrollment has also started to improve, but there is still a lack of completion of higher education (senior high school). Challenge: While many attend junior secondary school, the distribution of education quality at the senior secondary level and disparities between regions remain. Examples of Regions in this Cluster:

Fakfak: Has some relatively good schools but limited access to education and logistical challenges hinder equitable distribution of education.

Mimika: A region with fairly good education conditions, although there are still challenges in the distribution and quality of education.

Cluster 3 - Low (Major Challenges in Education Access)

Distinctive Features:

Low SPR at 7-12 years old (Z1), indicating that many young children in this cluster do not attend school or enter school late. Low GER for Senior Secondary School (Z6) and low NER for Senior Secondary School (Z9), reflecting that participation rates in senior secondary education are very limited. Average Years of Schooling (Z10) is also lower, reflecting uneven levels of education or that many only go to primary or junior secondary school. Challenges: Many external factors such as limited infrastructure, transportation access, and geographical conditions affect the low level of education. Examples of Regions in this Cluster:

Puncak Jaya: A mountainous area that is difficult to access and has major challenges in providing adequate education facilities, as well as high dropout rates.

Yahukimo: Also faces difficulties in accessing basic education due to limited infrastructure and lack of schools in remote areas.

From the above description, it is clear that there is a difference between areas with high and low levels of education.

IV. CONCLUSION

Some conclusions that can be drawn from the data and the results of the discussion are the formation of the three best clusters of clustering districts/cities on the island of Papua based on education indicators, consisting of:

1) Members of Cluster 1: Merauke, Jayawijaya, Jaya Pura, Nabire, Yapen Islands, Biak Nufor, Mimika, Boven Digoel, Sarmi, Keerom, Waropen, Supiori, Mamberamo Raya districts. Jaya Pura City, Fakfak, Kaiman, Teluk Wondama, Teluk Bintuni, Manokwari, South Sorong, Sorong, Raja Ampat, Tambrauw, Maybrat, South Manokwari, Arfak Mountains, Sorong City are part of Cluster1. The population in this cluster has a high level of education.
2) Members of Cluster 2: Mappi, Asmat, Memberamo Tengah, Dogiyai, Yalimo, Lani Jaya, Panilai districts. The population in this category has a medium level of education.
3) Members of Cluster 3: Nduga, Puncak, Intan Jaya, Yahukimo, Tolikura, Deiyai, Puncak Jaya, Pegunungan Bintang. The population in this cluster is categorized as having a low level of education.

In this clustering, the average value of the variables for each level of education in the cluster is still not the same. This means that the vision of education equity has not yet been achieved. To provide a clearer understanding of the regional clustering based on educational indicators, **Figure 3.** presents a spatial visualization of the cluster results for regencies/municipalities in Papua Island. The map illustrates the geographical distribution of the three clusters generated using the Ward's method, where each color represents a distinct cluster with similar educational characteristics. This spatial representation helps to identify regional disparities and patterns that may not be apparent from numerical data alone, and thus serves as a valuable tool for interpreting the clustering outcome in a policy-relevant contex.

BIBLIOGRAPHY

[1] A. F. Dewi and K. Ahadiyah, "Agglomerative Hierarchy Clustering Pada Penentuan Kelompok Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Pendidikan," *Zeta-Math J.*, vol. 7, no. 2, pp. 57–63, 2022.

[2] Badan Pusat Statistik Papua, "Indeks Pendidikan Provinsi Papua Tahun 2023," Badan Pus. Stat. Propinsi Papua, vol. 12, 2024.

[3] M. S. Ummah, "Pembangunan Bidang Pendidikan di Provinsi Papua 2023," Sustain., vol. 11, no. 1, pp. 1–14, 2019, [Online]. Available:

[4] Supranto, Analisis Multivariat, 1st ed. Jakarta: Cipta, Rineka, 2004.

[5] S. Santoso, Statistik multivariat. Elex media komputindo, 2010.

[6] Purnomo et al., Analisis Data Multivariat. 2022.

[7] B. Everitt, Cluster analysis, vol. 14, no. 1. 1980. doi: 10.1007/BF00154794.

[8] K. R. Subbaswamy, "Brillouin Scattering From Thermal Fluctuations in Superionic Conductors.," Solid State Commun., vol. 21, no. 4, pp. 371–372, 1977, doi: 10.1016/0038-1098(77)91248-0.

[9] D. Ls, Y. A. Lesnussa, M. W. Talakua, and M. Y. Matdoan, "Analisis Klaster untuk Pengelompokkan Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Pendidikan dengan Menggunakan Metode Ward," J. Stat. dan Apl., vol. 5, no. 1, pp. 51–60, 2021.

[10] R. Ramadani and A. Salma, "Metode Average Linkage Dan Ward Dalam Pengelom Pokan Kesejahteraan Sumatera Barat Tahun 2021," J. Math. UNP, vol. 7, no. 3, pp. 11–24, 2022.

[11] J. H. Ward, "Hierarchical Grouping to Optimize an Objective Function," J*ournal of the American Statistical Association*, vol. 58, no. 301. pp. 236–244, 1963. doi: 10.1080/01621459.1963.10500845.

[12] D. T. Larose, "An introduction to data mining," Trad. Adapt. Thierry Vallaud, p. 23, 2005.

[13] A. C.Rencher, Methods of Multivariate Analysis, vol. 11, no. 1. 2004. [Online]. Available: https://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf

[14] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, "An extensive comparative study of cluster validity indices," Pattern Recognit., vol. 46, no. 1, pp. 243–256, 2013, doi: https://doi.org/10.1016/j.patcog.2012.07.021.

[15] J. C. Dunn†, "Well-Separated Clusters and Optimal Fuzzy Partitions," *J. Cybern.*, vol. 4, no. 1, pp. 95–104, Jan. 1974, doi: 10.1080/01969727408546059

[16] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: Wiley, 2009.

[17] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*. Prentice-Hall, 1988.

[18] D. Yim and K. T. Ramdeen, "Hierarchical cluster analysis: Comparison of three linkage measures and application to psychological data," *The Quantitative Methods for Psychology*, vol. 11, no. 1, pp. 8–21, 2015

[19] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognition*, vol. 46, no. 1, pp. 243–256, Jan. 2013.