JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i3.5865

APPLYING K-MEANS CLUSTERING FOR GROUPING PAPUA'S DISTRICTS BASED ON POVERTY INDICATORS ANALYSIS

Yusriana Chusna Fadilah[1]; Asrul Sani[2*]; Andrianingsih Andrianingsih[3]

Study Program of Information Systems[1, 3], Study Program of Information Technology[2]
Universitas Nasional, Jakarta, Indonesia[1, 2, 3]
https://www.unas.ac.id[1, 2, 3]
yuschusfa@gmail.com[1], asrulsani@civitas.unas.ac.id[2*], andrianingsih@civitas.unas.ac.id[3]

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *In the context of Indonesia's resource-rich development, poverty remains a major challenge, especially in Papua Province which has the highest poverty rate. Although Papua is rich in resources such as minerals, tropical forests, and biodiversity, challenges such as economic inequality, lack of infrastructure, and social conflict hinder economic and social progress. This research aims to implement the K-Means Clustering algorithm to cluster districts/cities in Papua based on poverty indicators, including the percentage of poor people, poverty line, average years of schooling, human development index, poverty depth index, poverty severity index, unemployment rate, and per capita expenditure. The research methodology includes data collection from the Central Statistical Agency (BPS), data processing through cleaning and transformation stages, and application of K-Means Clustering to determine the optimal cluster using the elbow method and silhouette score. The results show that the districts/cities in Papua can be grouped into two main clusters: C0, which indicates high poverty rates and C1, which indicates low poverty rates. This research is expected to provide a strategic foundation for the government to design more focused and effective development policies in reducing poverty in Papua.*

Keywords: *elbow method, k-means clustering, papua, poverty indicators, silhouette method.*

Intisari— *Dalam konteks pembangunan Indonesia yang kaya akan sumber daya alam, kemiskinan tetap menjadi tantangan besar, terutama di Provinsi Papua yang memiliki tingkat kemiskinan tertinggi. Meskipun Papua kaya akan sumber daya seperti mineral, hutan tropis, dan keanekaragaman hayati, tantangan seperti ketimpangan ekonomi, kurangnya infrastruktur, dan konflik sosial menghambat kemajuan ekonomi dan sosial. Penelitian ini bertujuan untuk mengimplementasikan algoritma K-Means Clustering guna mengelompokkan Kabupaten/Kota di Papua berdasarkan indikator kemiskinan, termasuk persentase penduduk miskin, garis kemiskinan, rata-rata lama sekolah, indeks pembangunan manusia, indeks kedalaman kemiskinan, indeks keparahan kemiskinan, jumlah pengangguran, dan pengeluaran per kapita. Metodologi penelitian mencakup pengumpulan data dari Badan Pusat Statistik (BPS), pemrosesan data melalui tahap pembersihan dan transformasi, serta penerapan K-Means Clustering untuk menentukan cluster optimal menggunakan metode elbow dan skor silhouette. Hasil penelitian menunjukkan bahwa Kabupaten/Kota di Papua dapat dikelompokkan menjadi dua cluster utama: C0, yang menunjukkan tingkat kemiskinan tinggi dan C1, yang menunjukkan tingkat kemiskinan rendah. Penelitian ini diharapkan dapat memberikan landasan strategis bagi pemerintah untuk merancang kebijakan pembangunan yang lebih terfokus dan efektif dalam mengurangi kemiskinan di Papua.*

Kata Kunci: *metode elbow, k-means clustering, papua, indikator kemiskinan, metode silhouette.*

INTRODUCTION

Indonesia, as a country in the process of development, is rich in natural resources such as petroleum, natural gas, and fertile agricultural products. However, despite its huge economic potential, the country faces serious challenges in overcoming poverty [1]. Poverty in Indonesia is often defined as an individual or group's difficulty in meeting their basic needs, while the surrounding environment does not provide sufficient opportunities to improve welfare or escape from a dangerous situation [2]. One of the main causes of poverty is high economic inequality. Although rapid economic growth has been seen in recent years, the benefits have not been evenly distributed, causing the gap between the rich and poor to widen. In addition, the lack of infrastructure development, especially in rural and remote areas, exacerbates this problem. Limited access to basic services such as education, health, clean water and sanitation complicates efforts to improve the lives of residents in these areas. High population growth also contributes to the poverty problem, where despite the decline in birth rates, the large population remains a burden for the government in addressing poverty [3]. Lack of access to quality education and job training adds to the difficulties, as many residents do not have skills that match the needs of the region.

Papua Province, located at the eastern tip of Indonesia, is a region with a very high poverty rate. Although rich in natural resources such as precious minerals, tropical forests and biodiversity, geographical constraints such as wilderness and large rivers hinder the development of infrastructure that is essential to improving community welfare. In addition, Papua also faces major challenges related to social and political conflicts that impact economic instability and development. The problem of poverty in Papua is closely related to social and political factors and the lack of access to education, health and other basic services. Data from the Central Statistics Agency (BPS) in 2023 showed that the percentage of poor people in Papua exceeded 15%, far above the national average of 9.36% [4]. Poverty in Papua reflects economic inequality as well as deep social and infrastructure problems. TNP2K (National Team for the Acceleration of Poverty Reduction) utilizes clustering techniques to better understand and address poverty. This technique helps in identifying significant patterns of poverty and determining clusters of areas that require special interventions. Clustering, or data grouping, divides large data into groups with similar characteristics, providing a strong basis for decision-making and poverty reduction strategies [5].

The use of the K-Means Clustering algorithm in this study is grounded in its ability to uncover hidden patterns in complex, multivariate data that are difficult to interpret manually. Clustering techniques like K-Means are particularly effective in grouping data points with similar characteristics, providing actionable insights for decision-making. Previous research by Bahauddin et al. [2] demonstrated the success of this method in clustering Indonesian provinces based on poverty levels, enabling policymakers to design targeted development strategies. Similarly, Rahmadani et al. [6] applied K-Means Clustering to identify traffic accident-prone areas, showcasing the algorithm's versatility in analyzing diverse datasets. These findings emphasize the practical utility of clustering in addressing societal challenges. For Papua Province, with its varied poverty indicators, K-Means offers a robust analytical framework to segment districts or cities effectively. This approach enables a nuanced understanding of poverty distribution, paving the way for more focused and effective interventions tailored to local needs.

This research aims to implement the K-Means Clustering algorithm to group districts/cities in Papua Province based on poverty indicators, with the hope of identifying significant poverty patterns in each region. This analysis is expected to provide an in-depth insight into the variation of poverty characteristics in various districts/municipalities, so as to design development policies that are more focused and in accordance with the needs of each region. By using clustering techniques, this research aims to provide a strategic foundation for the government in designing more effective and efficient policies in poverty reduction in Papua. The study provides valuable insights but has yet to be fully implemented or disseminated to the government. Future work should focus on integrating clustering results into actionable government policies, conducting pilot programs in selected districts, and developing automated tools to monitor poverty dynamics continuously for more effective interventions and sustainable impact.

In research on poverty clustering, various methods and regions have been investigated to understand existing poverty patterns. Previous research used the K-Means Clustering method to cluster poverty in Java Island, which revealed significant differences between clusters with low and high poverty rates [7]. In addition, another study that also applied the K-Means Clustering method in the Maluku Islands and Papua found differences in poverty levels between these regions

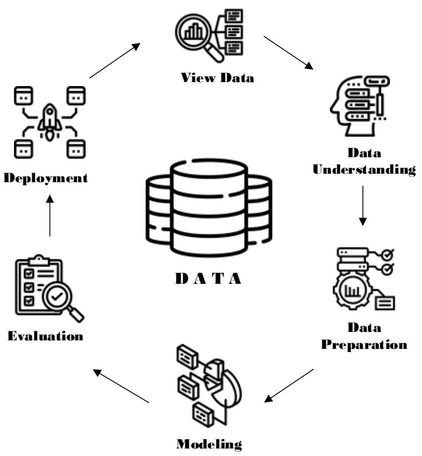
VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i3.5865

[8]. At the provincial level, poverty analysis in 34 Indonesian provinces identified three main clusters based on poverty levels: very poor, poor, and near poor [9]. In district/cities Papua, the Hierarchical Clustering method with the complete linkage technique produced three optimal clusters [5]. Meanwhile, poverty clustering in Banten Province using the K-Means Clustering method also showed three clusters with different poverty levels. These studies provide a diverse picture of the distribution and clustering of poverty in various regions, and offer different methods for poverty analysis [10].

Through the use of the K-Means Clustering algorithm [6][11], this research aims to provide a deeper insight into the pattern of poverty in Papua and formulate more effective intervention strategies according to the characteristics of each region.

MATERIALS AND METHODS

This research uses a quantitative approach with a focus on practical solutions to address poverty challenges in Papua Province. The main objective of this research is to cluster districts/cities in Papua based on poverty indicators using datasets from the Central Statistical Agency (BPS) from 2022 to 2023. The clustering process is carried out using the K-Means algorithm[12], with the determination of the optimal number of clusters using the Elbow and Silhouette methods [13]. These two methods were chosen to ensure accurate clustering results. The end result of this research includes the development of a web-based visualization tool that aims to assist the government in understanding and addressing poverty conditions. This section will explain the methodology in detail, including the stages of data collection, processing, and analysis and visualization that will be presented in Figure 1.



Source: (Ratner, 2017) [14]

Figure 1. K-Means Clustering Methodology

There are multiple steps to this method, namely:

View Data

In this research, the dataset used is "Poverty Data and Information" to assess the poverty rate of districts/cities in Papua Province from 2022 to 2023. The dataset was obtained from the official website of the Central Statistics Agency (BPS) under the titles "Poverty Data and Information in Indonesia 2022" and "Poverty Data and Information in Indonesia 2023". This initial stage involved checking the data type to ensure the relevance and availability of the data required for the subsequent poverty analysis.

Data Understanding

In this research, an in-depth understanding of the data is essential to categorize districts/cities based on poverty indicators in Papua Province. The dataset obtained from the official BPS website includes poverty indicators for 2022 and 2023, with a total of 232 data covering 8 poverty indicators from 29 districts/cities in Papua Province as shown in table 1. The indicators used in this study are:

Table 1. Indicators of Poverty

No	Data Type
1	Poverty Line
2	Average years of schooling
3	Human Development Index
4	Poverty Depth Index
5	Poverty Severity Index
6	Employments Rate
7	The Percentage of Poor People
8	Per Capita Expenditure

Source: (Research Results, 2024)

Data Preparation

This stage involves data preparation, which includes several steps:

1. Data Preprocessing

Data preprocessing in K-means clustering includes a series of steps before the data is fed into the K-means algorithm. The purpose of this step is to prepare the data to match the assumptions and requirements of the K-means algorithm and improve the clustering results. This stage includes data cleaning, handling outliers, missing values, duplicate data, and label encoding.

2. Data Transformation

The next step is to transform the data by creating new variables, handling non-numeric variables, deleting non-numeric variables, and scaling/standardizing numeric data. The transformed data is then integrated and transmitted according to the purpose of data mining processing.

Modeling

In the modeling stage, the model is defined and algorithms and data mining tools are implemented, such as Google Colab and Python programming language [15]. The data analysis method applied is K-Means Clustering on poverty indicators in Papua Province after the data transformation stage. The Elbow method is used to determine the optimal number of clusters in K-means [16]. The Elbow method displays a curve graph showing the optimal number of clusters, where the curve shows the sharpest angle or elbow.

The basic calculations for the K-Means algorithm include:

- Determining the value of k as the number of clusters to be formed.
- Determining the initial centroid for each cluster.
- Calculating the distance of objects to the centroid.
- The Euclidean distance formula is used:

$$d\,(x,y) = \sqrt{\sum_{i=1}^{n}(xi - yi)^2} \quad (1)$$

- Grouping objects into clusters based on the closest distance to the centroid.
- Determine and calculate a new centroid using the average of each member in the cluster:

$$C_k = \frac{1}{n_k}\sum d_i \quad (2)$$

Evaluation

The evaluation stage involves analyzing the clustering results, including the interpretation of the data mining model calculations that have been performed. The evaluation is done using information that is relevant to achieving the research objectives. Silhouette Score is used to assess the quality of clusters formed during modeling. This score measures how well the clusters are formed by evaluating the silhouette score for each k value and determining the k value that produces the highest silhouette score [17]. The purpose of this evaluation is to ensure that the applied model is in line with the research objectives.

Deployment

In the final stage, implementation involves preparing a report based on the results of applying data mining using K-means clustering. The research results will be visualized through a web-based interface using the Streamlit framework [18]. This report will provide insights for the government to prioritize and address poverty issues in districts/cities in Papua Province.

RESULTS AND DISCUSSION

The data analysis process began with the view data stage, data was obtained from the Central Statistics Agency (BPS) under the titles “Poverty Data and Information in Indonesia 2022” and “Poverty Data and Information in Indonesia 2023”, based on availability and suitability for poverty analysis. Then was put together in CSV (Comma Separated Value).

In the data understanding stage, researcher obtained detailed information about the dataset to understand the structure and characteristics of the data. We examined the details of the dataset that had been imported into Google Colab. This process uses Python libraries such as pandas, numPy, scikit-learn, and several other libraries [19].

Table 2. Data Types in Dataset

Column	Data Type
District/City	object
Poverty Line	float64
Average years of schooling	float64
Human Development Index	float64
Poverty Depth Index (P1)	float64
Poverty Severity Index (P2)	float64
Unemployment’s Rate	float64
The Percentage of Poor People	float64
Per Capita Expenditure	float64

Source: (Research Results, 2024)

Based on Table 2, shows the data type in each column of the dataset including float for 8 columns and object for 1 column.

In the first stage of the data preparation, data preprocessing is carried out which includes several important steps, namely checking for missing values, checking for duplicate data, and label encoding.

Table 3. Missing Value Checking

Column	Percentage (Null)
District/City	0
Poverty Line	0
Average years of schooling	0
Human Development Index	0
Poverty Depth Index (P1)	0
Poverty Severity Index (P2)	0
Employments Rate	0
The Percentage of Poor People	0
Per Capita Expenditure	0

Source: (Research Results, 2024)

Based on Table 3, shows that there are no missing values in the dataset.

Table 4. Data Duplicate Checking

Column	Duplicate Value
District/City	0
Poverty Line	0
Average years of schooling	0

VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i3.5865

Column	Duplicate Value
Human Development Index	0
Poverty Depth Index (P1)	0
Poverty Severity Index (P2)	0
Employments Rate	0
The Percentage of Poor People	0
Per Capita Expenditure	0

Source: (Research Results, 2024)

Based on Table 4, also shows that there is data duplicate in the dataset.

After ensuring that the data is free from missing values and duplicate data, the next step is to perform label encoding to facilitate the data analysis process. At this stage, the column "The Percentage of Poor Population" is moved to the second position in the data frame to adjust the column order according to the analysis needs.

In addition, the names of the poverty indicator columns were changed to simpler and easier to understand variables, such as changing "The Percentage of Poor People" to "X1", "Poverty Line" to "X2", and so on up to "Per Capita Expenditure" to "X8".

Table 5. Label Encoding

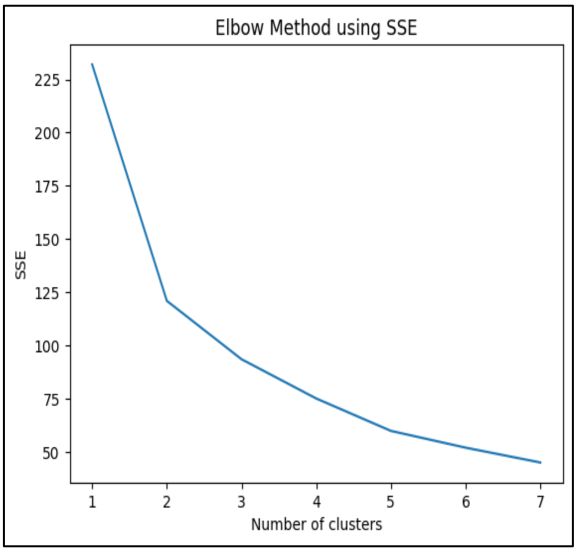
Column	Variable
The Percentage of Poor People	X1
Poverty Line	X2
Average years of schooling	X3
Human Development Index	X4
Poverty Depth Index (P1)	X5
Poverty Severity Index (P2)	X6
Employments Rate	X7
Per Capita Expenditure	X8

Source: (Research Results, 2024)

With the changes to Table 5, the final data frame presents the data in a cleaner and more structured format, facilitating the process of data analysis and interpretation in subsequent stages.

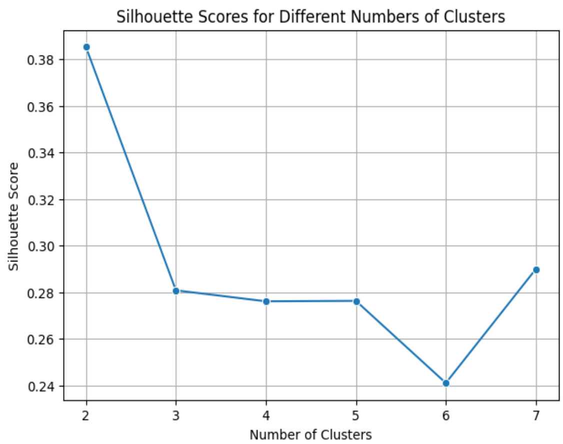
The data transformation stage involved scaling/standardizing numerical variables using StandardScaler from the sklearn library [20]. This process aims to change numerical variables to have the same scale, making it easier to form the optimal k in the modeling stage.

During the modeling stage, the elbow method was applied to identify the optimal number of clusters for K-Means Clustering based on Figure 2. This techniques help determine the most suitable cluster count by evaluating the sum of SSE for various values of k.



Source: (Research Results, 2024)
Figure 2. Elbow Method

In the evaluation stage, an assessment of the quality of the clusters formed from the previous modeling stage is carried out. The purpose of this evaluation is to ensure that the modeling applied is in accordance with the research and achieves the objectives set.



Source: (Research Results, 2024)
Figure 3. Silhouette Score

Table 6. Silhouette Score

Total Cluster	Silhouette Score
2	0.385370084101304
3	0.2808225787836018
4	0.27612178951588195
5	0.2763209017085812
6	0.24104738151026545
7	0.28997191985374116

Source: (Research Results, 2024)

The use of two methods in the K-Means clustering algorithm, namely elbow method (Figure 2) and silhouette score (Figure 3 and Table 6), shows that the optimal number of clusters is two clusters. In this process, the K-Means model is initialized with two clusters and trained using the scaled data. The model divides the data into two clusters, and the cluster labels are named C0 and C1 to distinguish between the two clusters. Thus, each data will have a C0 or C1 label according to the cluster determined by the K-Means model.

Finally, in the deployment stage, the cluster labels indicating the division of district groups were combined with the original data to provide a clearer and more informative context. This merging aims to make the resulting visualization present useful and easy-to-understand information. Using the Streamlit framework, the clustered data can be visualized interactively, aiming to explore and understand the distribution and characteristics of each cluster in more depth. This step not only facilitates the interpretation of the clustering results, but also improves the understanding of the patterns and relationships between variables in the analysis of poverty in Papua Province.

On the first page, there is a Data Overview that allows you to upload the dataset in CSV format. Next, it will display the contents and information of the dataset including checking for missing values, duplicate data, coding labels and transformations on the dataset.

Furthermore, the Clustering page contains information about K-Means theory including understanding, objectives, advantages and disadvantages. Next, it will enter the cluster calculation using two methods, namely elbow and silhouette, the results of which can be customized with features, namely in the form of graphs or scores. From the cluster calculation process using these two methods, it will display a table of clustering results, the number of each cluster and the average of each cluster.

Then the Visualizations page will display the poverty data table that has been clustered in tabular form. Furthermore, it will display a visualization in the form of a scatter plot to determine the characteristics of each cluster based on poverty indicators.

Finally, on the Strategy page, there are recommendations that can be implemented or updated based on the characteristics of each cluster. Recommendations here are in the form of existing and ongoing government programs or recommendations based on research analysis.

Determine K number of cluster centers randomly. In this first experiment, two random data are determined as the initial center point (centroid) for calculating the distance of all groups to be formed with the following details:

1. Number of clusters = 2 (taken randomly)
2. Number of data = 27
3. Number of attributes = 8

The random data can be seen in the following table:

Table 7. Centroid Iteration 1

Districts/Cities	Lanny Jaya	Kota Jayapura
X1	1,011133	-1,83031
X2	-0,32689	2,728522
X3	-0,936	1,759856
X4	-0,82137	2,037062
X5	2,268493	2,037062
X6	2,96143	-0,97976
X7	-1,38799	1,417614
X8	-1,00227	3,19063
Cluster	C0	C1

Source: (Research Results, 2024)

Based on Table 7, it can be seen that each cluster has the nearest distance seen in Table 8.

Table 8. Nearest Distance Iteration 1

C0	C1	Nearest Distance
7,392381	5,922813	5,922813
2,692909	4,645648	2,692909
7,435072	6,279172	6,279172
5,314926	4,5847	4,5847
5,38048	4,850957	4,850957
6,238906	4,903196	4,903196
5,141726	5,606165	5,141726
2,547368	6,25537	2,547368
7,4103	6,851669	6,851669
5,804982	4,934783	4,934783
6,134795	5,895851	5,895851
5,44581	5,775062	5,44581
4,244137	5,533674	4,244137
4,188493	6,205485	4,188493
4,440308	5,730834	4,440308
7,020983	6,319221	6,319221
6,851495	6,302611	6,302611
5,34168	5,281664	5,281664
3,992114	4,835772	3,992114
5,175804	6,47653	5,175804
3,183913	6,889518	3,183913
0	7,560318	0
3,392871	5,676661	3,392871
1,824084	6,337702	1,824084
3,320929	6,559687	3,320929
4,491513	5,382519	4,491513
2,926757	6,182456	2,926757
4,663719	6,320169	4,663719
9,321701	8,936184	8,936184

Source: (Research Results, 2024)

Calculate the distance of each data with each cluster center using the Euclidean Distance equation.

Table 9. Results Euclidean Distance Iteration 1

District/Cities	C0	C1
Merauke	7,392381	5,922813
Jayawijaya	2,692909	4,645648
Jayapura	7,435072	6,279172
Nabire	5,314926	4,5847

VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i3.5865

District/Cities	C0	C1
Kepulauan Yapen	5,38048	4,850957
Biak Numfor	6,238906	4,903196
Paniai	5,141726	5,606165
Puncak Jaya	2,547368	6,25537
Mimika	7,4103	6,851669
Boven Digoel	5,804982	4,934783
Mappi	6,134795	5,895851
Asmat	5,44581	5,775062
Yahukimo	4,244137	5,533674
Peg. Bintang	4,188493	6,205485
Tolikara	4,440308	5,730834
Sarmi	7,020983	6,319221
Keerom	6,851495	6,302611
Waropen	5,34168	5,281664
Supriori	3,992114	4,835772
Mamberamo Raya	5,175804	6,47653
Nduga	3,183913	6,889518
Lanny Jaya	0	7,560318
Mamberamo Tengah	3,392871	5,676661
Yalimo	1,824084	6,337702
Puncak	3,320929	6,559687
Dogiyai	4,491513	5,382519
Intan Jaya	2,926757	6,182456
Deiyai	4,663719	6,320169
Kota Jayapura	9,321701	8,936184

Source: (Research Results, 2024)

Based on Table 9, members are selected from the smallest of the two clusters with the following explanation:

1. The results of iteration 1 show that if the smallest value in the C0 section is included as a member of C0, which is 17 data district/cities.
2. The results of iteration 1 show that if the smallest value in the C1 section is included as a member of C1, which is 12 data district/cities.

Next, iteration 2 is carried out to ensure that the results of iteration 1 have no change in members in the cluster.

Determine the new centroid position for iteration 2 by calculating the average of the data in the centroid in iteration 1. The results of these calculations can be seen in Table 10 as follows:

Table 10. Centroid Iteration 2

Districts/Cities	C0	C1
X1	0,693031	-0,98179
X2	-0,24141	0,3420011,
X3	-0,7292	03304
X4	-0,7011	0,99322
X5	0,518057	-0,73391
X6	0,40271	-0,57051
X7	-54591	0,773371
X8	-0,64079	0,907788

Source: (Research Results, 2024)

Then calculate the distance of each data with each cluster center. The calculation is the same as the calculation stage iteration 1.

The results of these calculations are shown in Table 11 below:

Table 11. Results Euclidean Distance Iteration 2

District/Cities	C0	C1
Merauke	4,958364	2,346987
Jayawijaya	1,876916	4,129445
Jayapura	5,004541	1,151137
Nabire	3,464095	1,571999
Kepulauan Yapen	3,160976	1,407886
Biak Numfor	3,964743	0,955299
Paniai	2,038312	3,319416
Puncak Jaya	1,729258	4,616416
Mimika	5,395878	2,266759
Boven Digoel	3,19927	1,255956
Mappi	3,272546	2,590694
Asmat	2,46822	3,210449
Yahukimo	1,323755	4,162682
Peg. Bintang	1,268962	3,708375
Tolikara	1,591886	4,127274
Sarmi	4,282549	1,641798
Keerom	4,25965	0,852079
Waropen	3,129932	1,984076
Supriori	3,166879	4,077697
Mamberamo Raya	2,607307	3,073375
Nduga	2,41556	6,171615
Lanny Jaya	3,258309	6,434117
Mamberamo Tengah	1,160094	4,709548
Yalimo	2,27487	5,471712
Puncak	1,714051	5,041228
Dogiyai	1,583663	3,391382
Intan Jaya	1,580091	4,828254
Deiyai	1,856722	4,038246
Kota Jayapura	7,266385	3,768474

Source: (Research Results, 2024)

Based on Table 11, members are selected from the smallest of the two clusters with the following explanation:

1. The results of iteration 2 show that if the smallest value in the C0 section is included as a member of C0, which is 17 data district/cities.
2. The results of iteration 2 show that if the smallest value in the C1 section is included as a member of C1, which is 12 data district/cities.

The calculation results are obtained after two iterations which show that the members in the cluster have not changed and produce two clusters.

The following is the total number of members in each cluster show in Table 12.

Table 12. Members of the Cluster

Cluster	Total of Members	Cluster Members
C0	17	Jayawijaya, Paniai, Puncak Jaya, Asmat, Yahukimo, Pegunungan Bintang, Tolikara, Supiori, Mamberamo Raya, Nduga, Lanny Jaya, Mamberamo Tengah, Yalimo, Puncak, Dogiyai, Intan Jaya, Deiyai
C1	12	Merauke, Jayapura, Nabire, Kepulauan Yapen, Biak Numfor, Mimika, Boven Digoel, Mappi, Sarmi, Keerom, Waropen, Kota Jayapura

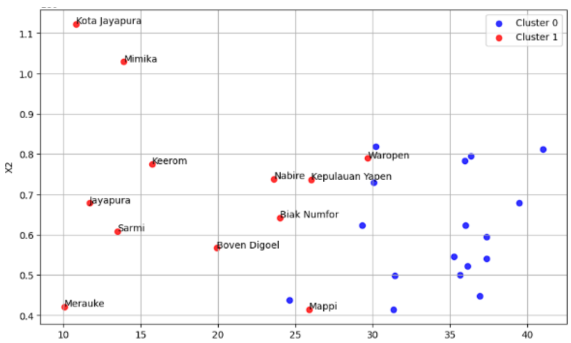
Source: (Research Results, 2024)

In this visualizations section, the final results of clustering using the K-Means method to group districts/cities in Papua Province based on poverty indicators will be presented through various displays. This visualization process is carried out using Google Colab to facilitate analysis and provide deeper insight into the poverty conditions in the region.

This visualization aims to present clear and directed information about the division of districts/municipalities in the two clusters produced. By utilizing visualization using scatter plots, it can illustrate the differences in characteristics between each cluster.

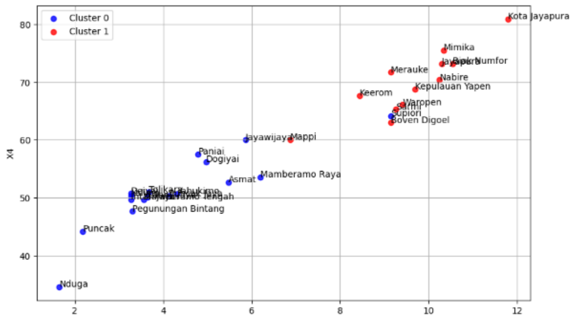
Scatter plots are designed to display the interaction between two variables at once, providing a clearer perspective on how these variables influence each other in the context of clustering. In this visualization, various combinations of relevant poverty variables will be analyzed, including The Percentage of Poor People with Poverty Line, Average Years of Schooling with Human Development Index, Poverty Depth Index (P1) with Poverty Severity Index (P2), and Unemployment Rate with Per Capita Expenditure.

Through the scatter plot visualization, we can see the patters that exist between the different poverty indicators in each cluster. The following some of the variable combinations chosen for the scatter plot visualization in four figures, each of which explains the pattern between the two variables.



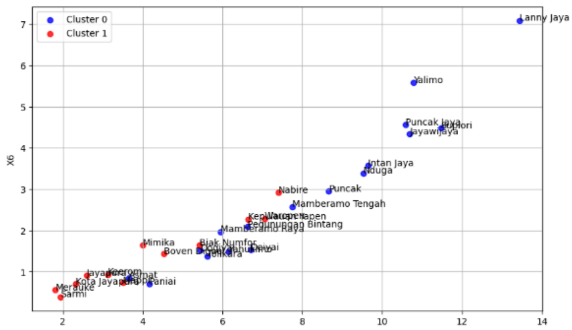
Source: (Research Results, 2024)
Figure 4. Scatter Plot: The Percentage of Poor People vs Poverty Line

Based on Figure 4, it can be seen that the two Districts/Cities clusters, C0 and C1, show significant differences in terms of poverty. The districts in C0 have a higher poverty line (between IDR 400,000 and 700,000 per capita per month) and a higher percentage of poor people compared to C1, which has a lower poverty line (between IDR 300,000 and 500,000 per capita per month). C0 reflects areas with greater poverty challenges, while C1 includes areas with lower poverty levels. The percentage of poor people and the poverty line are interrelated, where a higher poverty line often corresponds to a larger proportion of poor people. These results suggest that C0 faces more severe poverty challenges and requires more in-depth intervention strategies compared to C1.



Source: (Research Results, 2024)
Figure 5. Scatter Plot: Average Years of Schooling vs Human Development Index

Based on Figure 5, districts in C0 show lower average years of schooling and Human Development Index (HDI) compared to C1. The average years of schooling in C0 ranges from 2 to 9 years, with a HDI between 35 to 64, reflecting lower levels of education and human development. In contrast, C1 has a higher average year of schooling, between 7 to 12 years, and a HDI between 60 to 81, indicating better levels of education and human development. This comparison highlights the importance of education in improving quality of life, with C0 requiring improvements in education and quality of life, while C1 shows that better education contributes to a higher quality of life.



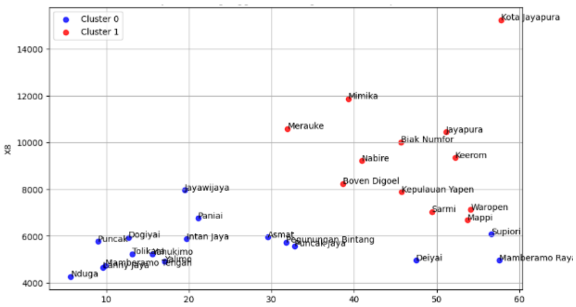
Source: (Research Results, 2024)
Figure 6. Scatter Plot: Poverty Depth Index vs Poverty Severity Index

Based on Figure 6, the districts in C0 have a Poverty Depth Index (P1) between 3.66 and 13.44 and a Poverty Severity Index (P2) between 0.705 and 7.085, indicating that the poor in C0 are far below the poverty line and experience higher

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)
VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i3.5865

income inequality. In contrast, C1 has a P1 between 1.815 and 7.400 and a P2 between 0.390 and 2.915, indicating that the poor in C1 are closer to the poverty line and experience lower income inequality. This difference emphasizes the need for different approaches in addressing poverty in each cluster.



Source: (Research Results, 2024)
Figure 7. Scatter Plot: Unemployment Rate vs Per Capita Expenditure

Based on Figure 7 there is a notable difference between C0 and C1 in terms of unemployment and per capita expenditure. The districts in C0 have unemployment between 5,690 and 57,545 and per capita expenditure between 4,271 and 7,962 rupiah, indicating smaller economic challenges. In contrast, C1 has unemployment between 31,925 and 57,785 and per capita expenditure between 6,685.5 and 15,230.5 rupiah, reflecting greater economic challenges. This analysis shows that C0 faces smaller unemployment and expenditure challenges than C1, which requires a more complex economic approach.

In this stage, there will be an in-depth explanation of the characteristics of each cluster resulting from the clustering analysis. This explanation aims to provide a clear insight into the differences and similarities between the clusters formed, thus providing a clear picture of the social and economic conditions of the districts/cities that influence poverty in Papua Province.

Table 13. Characteristics Cluster

Cluster	Districts/Cities	Characteristics	Category
C0	Jayawijaya, Paniai, Puncak Jaya, Asmat, Yahukimo, Pegunungan Bintang, Tolikara, Supiori, Mamberamo Raya, Nduga, Lanny Jaya, Mamberamo Tengah, Yalimo, Puncak, Dogiyai, Intan Jaya, dan Deiyai	High percentage of poor, high poverty line, low average years of schooling, low HDI, high P1 & P2, low unemployment and low per capita expenditure.	High Poverty Level
C1	Merauke, Jayapura, Nabire, Kepulauan Yapen, Biak Numfor, Mimika, Boven Digoel, Mappi, Sarmi, Keerom, Waropen, dan Kota Jayapura	Low percentage of poor, low poverty line, high average years of schooling, high HDI, low P1 & P2, high unemployment and high per capita expenditure.	Low Poverty Level

Source: (Research Results, 2024)

CONCLUSION

This research successfully implemented K-Means Clustering to group districts/cities in Papua Province based on poverty indicators, resulting in two clusters: C0 and C1. The Elbow and Silhouette Score methods show that the two clusters are optimal. Cluster C0, which is categorized as having a high poverty rate, includes 17 districts/cities: Jayawijaya, Paniai, Puncak Jaya, Asmat, Yahukimo, Pegunungan Bintang, Tolikara, Supiori, Mamberamo Raya, Nduga, Lanny Jaya, Mamberamo Tengah, Yalimo, Puncak, Dogiyai, Intan Jaya, and Deiyai. In contrast, Cluster C1, which is categorized as a low poverty rate, includes 12 districts/cities, namely Merauke, Jayapura, Nabire, Yapen Islands, Biak Numfor, Mimika, Boven Digoel, Mappi, Sarmi, Keerom, Waropen, and Kota Jayapura.

The results of this study provide a strategic foundation for the government to design development policies that are more focused and tailored to overcome poverty in Papua Province. By identifying two distinct clusters with varying poverty levels, this analysis highlights the unique challenges faced by each group, enabling more effective and efficient policy-making. These findings can be leveraged to develop targeted programs addressing critical issues such as education, infrastructure development, and economic inequality. Such initiatives will not only tackle immediate poverty concerns but also promote long-term sustainable development. By aligning these strategies with the specific needs of each cluster, the government can ensure more equitable progress, fostering improved welfare and reducing disparities across Papua Province.

REFERENCE

[1] L. S. Hasibuan, P. Hariani, and A. S. Pratiwi, "Poverty development post fiscal decentralization policy in Otsus fund receiving province in Indonesia (Case study: Province of Papua, West Papua, and Aceh)," *International Journal of Economic, Technology and Social Sciences (Injects),* vol. 2, no. 2, pp. 562-574, 2021. DOI: https://doi.org/10.53695/injects.v2i2.598

[2] A. Bahauddin, A. Fatmawati, and F. P. Sari, "Analisis Clustering Provinsi Di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means," *Jurnal Manajemen Informatika Dan Sistem Informasi,* vol. 4, no. 1, pp. 1-8, 2021. DOI: https://doi.org/10.36595/misi.v4i1.216

[3] E. Andriaswati and S. Utami, "Determinants of poverty rates in Papua province in 2011-2019," *Efficient: Indonesian Journal of Development Economics,* vol. 5, no. 1, pp. 1453-1467, 2022. DOI: https://doi.org/10.15294/efficient.v5i1.50928

[4] A. W. F. Amin, "The Dynamic of Education Outcome in Decentralized Indonesia," *Indonesia at The Crossroads: Transformation and Challenges,* p. 149, 2023.

[5] F. B. Putra, A. T. R. Dani, and S. Wigantono, "Penerapan Algoritma Hierarchical Clustering Dalam Pengelompokkan Kabupaten/Kota Di Papua Berdasarkan Indikator Kemiskinan," *MAp (Mathematics and Applications) Journal,* vol. 5, no. 2, pp. 118-126, 2023. DOI: https://doi.org/10.15548/map.v5i2.7025

[6] N. Rahmadani, E. Rahayu, and A. Lestari, "K-Means Clustering Areas Prone To Traffic Accidents in Asahan Regency," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer),* vol. 6, no. 2, pp. 181-186, 2021.DOI: https://doi.org/10.33480/jitk.v6i2.1519

[7] I. Agustina and R. D. Dana, "Implementasi K-Means Clustering Untuk Pengelompokan Tingkat Kemiskinan Di Pulau Jawa Menggunakan Bahasa Pemrograman Python," *JATI (Jurnal Mahasiswa Teknik Informatika),* vol. 8, no. 1, pp. 945-951, 2024. DOI: https://doi.org/10.36040/jati.v8i1.8858

[8] M. Matdoan, L. Igo, R. Rumeon, R. Fadhilah, and N. Laamena, "Penerapan Algoritma K-Means Untuk Klusterisasi Kabupaten/Kota Berdasarkan Tingkat Kemiskinan di Kepulauan Maluku dan Papua," *Jurnal Sains Matematika dan Statistika,* vol. 10, no. 1, pp. 1-9. DOI: http://dx.doi.org/10.24014/jsms.v10i1.21260

[9] A. Khalif, A. N. Hasanah, M. H. Ridwan, and B. N. Sari, "Klasterisasi Tingkat Kemiskinan di Indonesia menggunakan Algoritma K-Means," *Generation Journal,* vol. 8, no. 1, pp. 54-62, 2024. DOI: https://doi.org/10.29407/gj.v8i1.21470

[10] A. Sani, A. Andrianingsih, S. Aisyah, and A. Taufik, "Dampak Kesiapan Organisasi terhadap Keberhasilan Cloud Computing di UMKM dengan Model Delone Dan Mclean," *Infotech: Journal of Technology Information,* vol. 10, no. 2, pp. 319-326, 2024. DOI: https://doi.org/10.37365/jti.v10i2.325

[11] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *Evolusi: Jurnal Sains dan Manajemen,* vol. 9, no. 1, 2021. DOI: https://doi.org/10.31294/evolusi.v9i1.10428

[12] R. Buaton and S. Solikhun, "A Comparative Evaluating Numerical Measure Variations in K-Medoids Clustering for Effective Data Grouping," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer),* vol. 10, no. 2, pp. 394-403, 2024. DOI: https://doi.org/10.33480/jitk.v10i2.5545

[13] C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, and J. Liu, "A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm," *EURASIP journal on wireless communications and networking,* vol. 2021, pp. 1-16, 2021. DOI: https://doi.org/10.1186/s13638-021-01910-w

[14] B. Ratner, B. Ratner, Ed. *Statistical and Machine Learning. Techniques for Better Predictive Modelling and Analysis of Big Data*. U.S.A: Taylor & Francis Group, 2017, p. 655.

[15] P. Naik, G. Naik, and M. Patil, "Conceptualizing Python in Google COLAB," *India: Shashwat Publication,* 2022.

[16] S. Saraswat, V. Agrohi, M. Kumar, M. Lamba, and R. Kaur, "Unveiling Consumer Segmentation: Harnessing K-means Clustering Using Elbow and Silhouette for Precise Targeting," in *International Conference on Computing and Communication Networks*, 2023, pp. 355-369: Springer. DOI: https://doi.org/10.1007/978-981-97-0892-5_28

VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i3.5865

[17] A. Punhani, N. Faujdar, K. K. Mishra, and M. Subramanian, "Binning-based silhouette approach to find the optimal cluster using K-means," *IEEE Access,* vol. 10, pp. 115025-115032, 2022. DOI: https://doi.org/10.1109/ACCESS.2022.3215568

[18] M. Khorasani, M. Abdou, and J. H. Fernández, *Web application development with streamlit: Develop and deploy secure and scalable web applications to the cloud using a pure Python framework*. Springer, 2022. DOI: https://doi.org/10.53695/injects.v2i2.598

[19] P. Lemenkova, "Processing oceanographic data by Python libraries NumPy, SciPy and Pandas," *Aquatic Research,* vol. 2, no. 2, pp. 73-91, 2019. DOI: https://doi.org/10.3153/AR19009

[20] A. Testas, "Support Vector Machine Classification with Pandas, Scikit-Learn, and PySpark," in *Distributed Machine Learning with PySpark: Migrating Effortlessly from Pandas and Scikit-Learn*: Springer, 2023, pp. 259-280. DOI : https://doi.org/10.1007/978-1-4842-9751-3_10