International Journal of Computer and Information System (IJCIS)
Peer Reviewed – International Journal
Vol : Vol. 06, Issue 01, February 2025
e-ISSN : 2745-9659
https://ijcis.net/index.php/ijcis/index

Identifying Regional Patterns of Poverty in Indonesia : a Clustering Approach Using K-Means

Sri wahyuni[1], Agustia Hananto[2], Baenil Huda[3], Fitria Apriani[4], Tukino[5]
[1,2,3,4,5]Information System
Buana Perjuangan University Karawang
Karawang, Indonesia
si18.sriwahyuni@mhs.ubpkarawang.ac.id[1], agustia.hananto@ubpkarawang.ac.id[2], baenil88@ubpkarawang.ac.id[3], fitria.apriani@ubpkarawang.ac.id[4], tukino@ubpkarawang.ac.id[5]

Abstract-Poverty in Indonesia remains a major challenge, with significant levels of inequality between provinces. This study applies the K-Means clustering method to analyze poverty distribution patterns in 38 provinces in Indonesia, using data on the percentage of poor people from 2010 to 2024. With this approach, provinces are grouped into three main clusters: low, medium, and high, based on the average poverty rate. The low cluster includes areas with poverty rates below 10%, while the medium and high clusters indicate poverty levels that require more specific policies. The evaluation results show a silhouette score of 0.613, indicating that this grouping is quite good but can still be improved. This study offers important insights to support more targeted and effective policies, especially in achieving Sustainable Development Goal (SDG) 1: Eradicating Poverty.

Keyword : *Poverty, Clustering, K-Means, Indonesia*

I. INTRODUCTION

Poverty is a complex issue involving multiple social, economic, and political dimensions (Kliuchnyk, 2022). In Indonesia, poverty remains one of the main challenges to sustainable development (Sumargo & Haida, 2020). Poverty rates in Indonesia vary widely across provinces, with some regions, such as Papua and Aceh, having much higher poverty rates than other regions (Erda et al., 2023). Despite a decline in poverty nationally, poverty disparities across provinces show significant patterns (Hidayat et al., 2022). Several provinces, especially in eastern Indonesia, still show much higher poverty rates compared to provinces in the west (BPS, 2022). This highlights the importance of a data-driven approach to understanding the varying distribution of poverty across Indonesia. One way to analyze these poverty patterns is to use clustering techniques, which can group provinces with similar poverty characteristics.

Research by (Foell & Pitzer, 2020) found that regional clustering can identify areas that require special policies, such as areas with limited infrastructure and low access to education. Spatial clustering has also been applied to identify pockets of poverty in developing countries, such as research by (Amarasinghe et al., 2020). In the Indonesian context, clustering poverty levels between provinces is important given the significant disparity between eastern and western Indonesia (Kurniasari & Oktavilia, 2023). Previous studies have focused more on descriptive analysis, while data mining-based approaches are still under-optimized (Islam & Khan, 2020).

Empirical studies have shown how clustering can be used to analyze poverty distribution. Research by (Sano & Nindito, 2020) used the K-Means method to cluster poverty levels in various provinces in Indonesia, showing that areas with similar geographic characteristics tend to have similar poverty levels. K-Means in analyzing the percentage of poor population dataset lies in its efficiency in processing simple numerical data like this (Yu et al., 2020). In addition, K-Means results are easy to interpret through centroids that represent the average characteristics of each cluster (Kim et al., 2020).

K-Means, as one of the most common methods, works by grouping data based on their proximity to predetermined cluster centers (Hasan et al., 2023). The application of clustering in poverty analysis provides valuable insights for both researchers and policymakers. Research by (Sano & Nindito, 2020) shows that, for example, by grouping provinces based on poverty levels and other related factors, the government can design more specific and effective policies according to the characteristics of each region. Research by (Poerwanto, 2023) reveals that the use of clustering methods for spatial analysis of poverty in Indonesia has been proven to help formulate more inclusive regional development policies. This study offers a unique contribution by applying the K-Means clustering method to identify regional patterns of poverty in Indonesia, which can then be used to support the achievement of Sustainable Development Goal (SDG) 1: Eradicating Poverty (Sano & Nindito, 2020).

Clustering is one of the main techniques in data mining used to group data into groups based on similar characteristics (Chaudhry et al., 2023). In socio-economic analysis, clustering allows the identification of hidden patterns or trends in large datasets (Irani et al., 2020). In data mining, grouping data into several groups (clusters) based on similar characteristics or patterns that are owned as an unsupervised learning method, clustering does not require previous data labels or categories, but rather relies on the structure of the data itself to form groups (Chander & Vijaya, 2021). The clustering process involves steps such as data preprocessing, selecting an appropriate clustering algorithm (e.g., K-Means, Hierarchical Clustering, or DBSCAN), and evaluating clustering results using metrics such as the Elbow Method, Silhouette Score, or Davies-Bouldin Index.

Clustering aims to minimize the distance between data in one cluster and maximize the distance between different clusters. In other words, data that are more similar to each other will be grouped together in one cluster, while data that is very different will be separated into different clusters (Gonzalez, 2020). For example, in the context of poverty research, we can group provinces in Indonesia based on similarities in their economic and social characteristics, such as poverty rates, unemployment, education, and others (Candra et al., 2024).

K-Means is a partition-based clustering algorithm that iteratively groups data into a predetermined number of clusters. This algorithm is known for its simplicity and efficiency in analyzing large amounts of data (Vrahatis et al., 2020). However, its sensitivity to outliers and dependence on the initial centroid value are drawbacks that need to be considered (He et al., 2020). Figure 1 below is a flowchart of the K-Means algorithm.

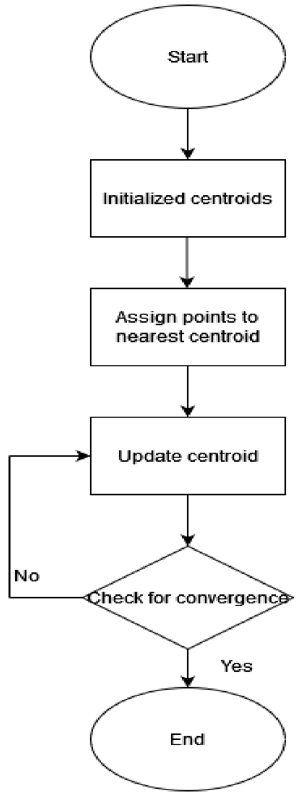

Figure 1. K-Means Flowchart

This study uses Google Colab as a cloud-based platform that allows researchers to write and run Python code in interactive notebooks without the need for local environment settings (Timpe, 2023). This service is integrated with Google Drive, allowing easy access, storage, and sharing of files. Colab also provides free access to GPUs and TPUs, which are very useful for accelerating computations that require high processing power, such as machine learning and deep learning. This platform supports various popular libraries such as TensorFlow, PyTorch, Pandas, and NumPy, which facilitate data analysis and model development (Li et al., 2024). Users can easily create, edit, and share notebooks with collaborators in real-time, making it an ideal choice for research teams or projects that require intensive collaboration. Users can also access data from Google Drive and upload files directly to the notebook. However, there are some drawbacks such as time limitations on GPU/TPU usage and storage limitations on Google Drive. Nevertheless, Google Colab remains a very useful tool for data analysis and machine learning development, especially for users who do not have powerful hardware or do not want to install software locally (Llerena-Izquierdo et al., 2024).

II. RESEARCH METHODS

In this study, the data analysis process was carried out using the K-Means clustering approach. To ensure valid and interpretable analysis results, systematic steps were followed from initial data processing to model performance evaluation. The flowchart can be seen in Figure 2. Below

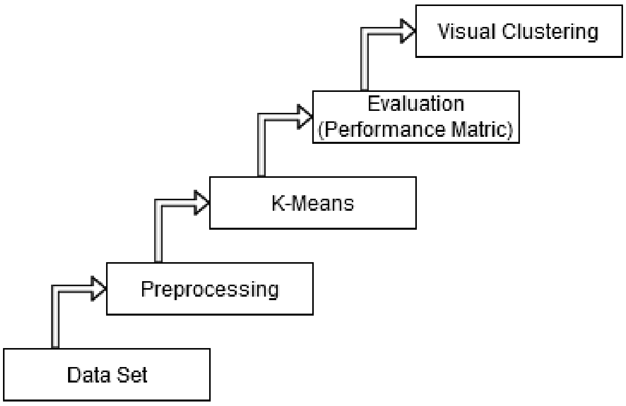

Figure 2. Research flow diagram

The image illustrates the overall stages carried out in this study. The process begins with a dataset, in the context of clustering in data mining, a dataset usually contains a set of raw data that has a number of features or attributes that represent the objects to be grouped which are the main input for analysis (Chaudhry et al., 2023). This dataset then goes through a preprocessing process, the preprocessing stage in clustering analysis is an important step to ensure that the data is clean, relevant, and ready to be used by the algorithm (Simon et al., 2020). This process begins by understanding the characteristics of the dataset through descriptive analysis to identify data distribution, missing values, and outliers. The next step is data cleaning, including imputation of missing values, removal of duplicate data, and correction of format errors (Winkler, 2020).

After preprocessing is complete, the data is analyzed using the K-Means clustering algorithm, the modeling stage in this study aims to group provinces in Indonesia based on poverty levels using this algorithm. The K-Means clustering algorithm works by dividing data into a number of groups (clusters) based on the proximity of the distance between data in multidimensional space. The number of clusters is determined through evaluation methods such as the Elbow Method to determine the optimal number of clusters by looking at the inertia graph (in this case, the total distance between the data and its centroid). A significant decrease in the graph indicates the right number of clusters. The point at which the decrease slows down is called the "elbow" and is an indication of the optimal number of clusters (Amit, 2021).

Silhouette Score is used to evaluate the K-Means method. Silhouette Score measures how well objects are grouped in their clusters, with a value between -1 and +1. The higher the Silhouette score, the better the clustering is. The optimal Silhouette score provides a strong indicator of the right number of clusters. After the clustering and evaluation process is complete, the results of the analysis will be visualized to provide a clear picture of the distribution of provinces based on poverty levels. For this visualization, a scatter plot is used that shows the grouping of provinces using different colors for each cluster. In the visualization of the clustering results, researchers will analyze the characteristics of each cluster to identify provinces with similar poverty patterns. This will provide insight into poverty disparities between provinces and allow researchers to provide more targeted policy recommendations. For example, provinces with high poverty rates and similar characteristics can be directed to receive more focused and area-based poverty alleviation policies.

III. RESULT AND ANALYSIS

The data was obtained from open data jabar obtained from the link https://opendata.jabarprov.go.id/id. This dataset has a tabular structure with a significant number of rows (570 rows), covering data for 38 provinces in Indonesia from 2010 to 2024. Which consists of variables ID, province code, province name, population percentage, unit, and year. Figure 3 below is a picture of the dataset.

```
          nama_provinsi  persentase_penduduk_miskin  tahun
0                  ACEH                       19.95   2010
1        SUMATERA UTARA                       11.36   2010
2        SUMATERA BARAT                        9.44   2010
3                  RIAU                       10.01   2010
4                 JAMBI                        8.40   2010
..                  ...                         ...    ...
565   PAPUA BARAT DAYA                        18.13   2024
566               PAPUA                       17.26   2024
567       PAPUA SELATAN                       17.44   2024
568        PAPUA TENGAH                       29.76   2024
569   PAPUA PEGUNUNGAN                        32.97   2024

[570 rows x 3 columns]
```

Figure 3. Dataset

International Journal of Computer and Information System (IJCIS)
Peer Reviewed – International Journal
Vol : Vol. 06, Issue 01, February 2025
e-ISSN : 2745-9659
https://ijcis.net/index.php/ijcis/index

This dataset contains information about the name of the province, the percentage of poor people, and the year of measurement. For the initial analysis, the first step taken is to select the relevant columns, namely the name_of_the_province, the percentage_of_poor_people, and the year. This selection aims to simplify the dataset so that it only contains information that is really needed for the calculation of the average poverty and clustering analysis. Next, the average poverty percentage is calculated for each province by grouping the data by province name. This calculation uses the mean function from the pandas library, which automatically calculates the average value of the percentage_of_poor_population column for each province. The result is a new data frame containing the name of the province and its average poverty value. To make it easier to interpret, the name of the column that shows the average poverty is changed to average_poverty.

The resulting data is then focused on the average_poverty_column, which is the main numeric attribute for analysis using K-Means. This column is chosen because clustering analysis requires numeric attributes to determine patterns or groups among data. However, in order for the clustering results to remain contextually interpretable, the data is recombined with the province name using the merge technique. This ensures that each clustering result still has the appropriate provincial information. Figure 4 below shows the average poverty per province.

	nama_provinsi	rata_rata_kemiskinan
0	ACEH	16.556000
1	BALI	4.375333
2	BANTEN	5.882667
3	BENGKULU	16.057333
4	DI YOGYAKARTA	13.374667
5	DKI JAKARTA	4.024000
6	GORONTALO	16.612667
7	JAMBI	8.022000
8	JAWA BARAT	8.742667
9	JAWA TENGAH	12.844667
10	JAWA TIMUR	11.822000
11	KALIMANTAN BARAT	7.698000
12	KALIMANTAN SELATAN	4.756000
13	KALIMANTAN TENGAH	5.687333
14	KALIMANTAN TIMUR	6.336000
15	KALIMANTAN UTARA	6.720667
16	KEPULAUAN BANGKA BELITUNG	5.126667
17	KEPULAUAN RIAU	6.281333
18	LAMPUNG	13.667333
19	MALUKU	18.862667
20	MALUKU UTARA	7.233333
21	NUSA TENGGARA BARAT	16.072000
22	NUSA TENGGARA TIMUR	20.864000
23	PAPUA	27.885333
24	PAPUA BARAT	24.385333
25	PAPUA BARAT DAYA	1.208667
26	PAPUA PEGUNUNGAN	2.198000
27	PAPUA SELATAN	1.162667
28	PAPUA TENGAH	1.984000
29	RIAU	7.690667
30	SULAWESI BARAT	11.997333
31	SULAWESI SELATAN	9.344000
32	SULAWESI TENGAH	13.954667
33	SULAWESI TENGGARA	12.518667
34	SULAWESI UTARA	8.019333
35	SUMATERA BARAT	7.064667
36	SUMATERA SELATAN	13.103333
37	SUMATERA UTARA	9.620667

Figure 4. Average poverty per province

To determine the optimal number of clusters in the K-Means Clustering process, the Elbow Method is used. This method helps identify the optimal point where the cluster division produces a good representation of the data without losing efficiency. In this step, the inertia value, which is the total squared distance between each data and its nearest centroid, is calculated for various numbers of clusters.

The range of the number of clusters tested ranges from 1 to 10. For each number of clusters, the K-Means algorithm is applied to the dataset, and the inertia value is calculated. This inertia indicates how well the data is clustered; the smaller the inertia value, the better the data is grouped into clusters. However, too many clusters can lead to overfitting, so a systematic approach is needed to determine the optimal number of clusters.

The results of the inertia calculation are then visualized in the form of an Elbow Curve graph. This graph plots the number of clusters on the x-axis and the inertia value on the y-axis. The optimal number of clusters is located at the point where the decrease in the inertia value begins to slow down significantly, forming an elbow-like angle. For example, if the graph shows a drastic decrease until the number of clusters is 3, and after that the decrease becomes more gentle, then the optimal number of clusters is 3.

By using the Elbow Method, we can ensure that the cluster division reflects patterns in the data in an efficient and meaningful way. In the context of a poverty dataset, the optimal number of clusters will provide a picture of a group of provinces with similar poverty level characteristics, which can be used for further analysis and more targeted policy planning. Figure 5 below is an Elbow curve.

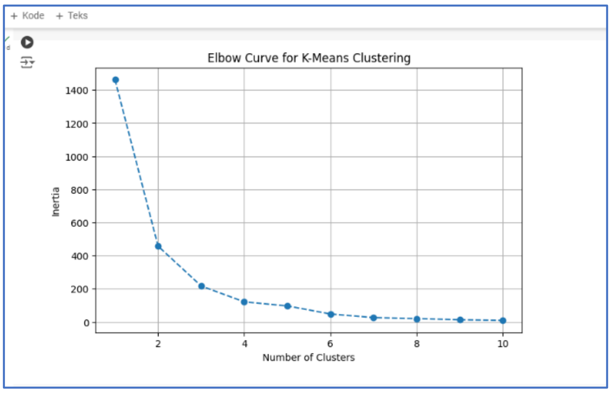

Figure 5. Elbow curve

In the next process, K-Means utilizes iteration through three clusters, which are determined by the cluster_id value. For each cluster, data is filtered from the dataset based on the corresponding cluster column. After the data is segmented, a scatter plot is created for each cluster by displaying the average poverty value as the X-axis. Each cluster is labeled with a name, such as "Cluster 0", "Cluster 1", or "Cluster 2", making it easier to identify in the plot. A simple visualization of the clustering results based on the average poverty level in various regions uses a Scatter plot. In this plot, the data is divided into three different clusters, where each cluster is represented by a collection of dots with different colors and labels. These dots are plotted on the horizontal axis (X-axis) according to the average poverty value, while the vertical axis (Y-axis) is intentionally set to zero to place all clusters on one horizontal line. This is done so that the differences between clusters are easier to see and are not disturbed by additional dimensions. Figure 6 below is a scatter plot of clustering.

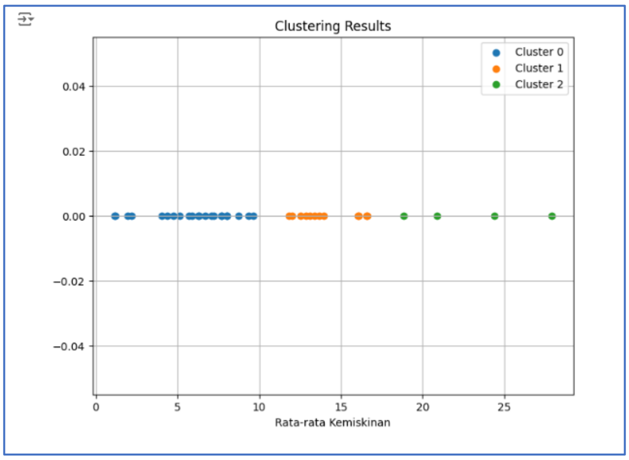

Figure 6. Scatter plot of clustering

The evaluation results of the K-Means analysis show that the silhouette score value obtained provides an overview of the quality of the clustering that has been done. The silhouette score value of 0.613 is quite good and the value of 0.613 is relatively close to 1, which indicates that most provinces are in the right cluster and quite far from other clusters. Although quite good, the value of 0.613 still leaves room for improvement. This means that there may be some provinces whose positions between clusters are unclear or are in clusters that are less appropriate. The calculation of the Silhouette score can be seen in Figure 7 below.

```
[7] from sklearn.metrics import silhouette_score

[8] score = silhouette_score(clustering_data, average_poverty['cluster'])
    print(f"Silhouette Score: {score}")

    Silhouette Score: 0.6126007535494638
```

Figure 7. Silhouette score

Analysis of poverty levels in Indonesia resulted in the division of provinces into three main clusters based on the average percentage of poor people. These clusters reflect the variation in socio-economic conditions in each province, ranging from areas with low poverty levels to areas with significant poverty challenges. The following explanation describes the characteristics of each cluster, including factors that influence poverty levels, as well as efforts that can be made to improve the welfare of the population in each group.

Cluster 1

Cluster 1 (Low) includes provinces in Indonesia that have an average poverty rate of below 10%. This group represents areas with a relatively small percentage of poor people compared to other provinces. Provinces in Cluster 1, such as Bali (4.375%), DKI Jakarta (4.024%), and Banten (5.883%), have shown success in maintaining local economic stability, providing adequate infrastructure, and good access to basic needs such as education, health, and employment. This shows that these areas are able to provide a better quality of life for their residents.

This achievement can be caused by several factors, such as advanced industrialization, a growing tourism sector, and effective local government policies in reducing poverty rates. With low poverty rates, provinces in Cluster 1 can focus more on efforts to improve the quality of life of the population and economic equality, rather than on large-scale poverty alleviation programs. This cluster shows the potential for economic sustainability and better welfare compared to other provinces in Indonesia.

nama_provinsi	rata_rata kemiskinan	cluster
BALI	4,375	Cluster 1 (Rendah)
BANTEN	5,883	Cluster 1 (Rendah)
DKI JAKARTA	4,024	Cluster 1 (Rendah)

JAMBI	8,022	Cluster 1 (Rendah)
JAWA BARAT	8,743	Cluster 1 (Rendah)
KALIMANTAN BARAT	7,698	Cluster 1 (Rendah)
KALIMANTAN SELATAN	4,756	Cluster 1 (Rendah)
KALIMANTAN TENGAH	5,687	Cluster 1 (Rendah)
KALIMANTAN TIMUR	6,226	Cluster 1 (Rendah)
KALIMANTAN UTARA	6,721	Cluster 1 (Rendah)
KEPULAUAN BANGKA BELITUNG	5,127	Cluster 1 (Rendah)
KEPULAUAN RIAU	6,281	Cluster 1 (Rendah)
MALUKU UTARA	7,233	Cluster 1 (Rendah)
PAPUA BARAT DAYA	1,209	Cluster 1 (Rendah)
PAPUA PEGUNUNGAN	2,198	Cluster 1 (Rendah)
PAPUA SELATAN	1,163	Cluster 1 (Rendah)
PAPUA TENGAH	1,984	Cluster 1 (Rendah)
RIAU	7,691	Cluster 1 (Rendah)
SULAWESI SELATAN	9,344	Cluster 1 (Rendah)
SULAWESI UTARA	8,019	Cluster 1 (Rendah)
SUMATERA BARAT	7,065	Cluster 1 (Rendah)
SUMATERA UTARA	9,621	Cluster 1 (Rendah)

Cluster 2

Based on the results of poverty data analysis, the provinces included in Cluster 2 (moderate category) have varying average poverty rates, but are still within a fairly significant range. The province with the highest poverty rate in this cluster is Maluku with an average of 18.863%, while the lowest is East Java with an average of 11.822%.

Other provinces included in this cluster include Aceh with an average poverty rate of 16.556%, followed by Gorontalo at 16.613%, and West Nusa Tenggara with a value of 16.072%. In addition, Bengkulu showed an average poverty rate of 16.057%, while Central Sulawesi and Southeast Sulawesi recorded values of 13.955% and 12.519% respectively. The province of DI Yogyakarta has an average poverty rate of 13.375%, slightly lower than Lampung which is at 13.667%.

For provinces in Java, apart from East Java, Central Java recorded an average poverty rate of 12.849%, slightly higher than Southeast Sulawesi. This cluster as a whole represents a group of provinces with moderate poverty rates, indicating the need for more focused attention and policy efforts to reduce poverty rates in these areas.

nama_provinsi	rata_rata_kemiskinan	Cluster
ACEH	16,556	Cluster 2 (Sedang)
BENGKULU	16,057	Cluster 2 (Sedang)
DI YOGYAKARTA	13,375	Cluster 2 (Sedang)
GORONTALO	16,613	Cluster 2 (Sedang)
JAWA TENGAH	12,849	Cluster 2 (Sedang)
JAWA TIMUR	11,822	Cluster 2 (Sedang)
LAMPUNG	13,667	Cluster 2 (Sedang)
MALUKU	18,863	Cluster 2 (Sedang)
NUSA TENGGARA BARAT	16,072	Cluster 2 (Sedang)
SULAWESI TENGAH	13,955	Cluster 2 (Sedang)
SULAWESI TENGGARA	12,519	Cluster 2 (Sedang)

Cluster 3

The provinces included in cluster 3 with a high poverty rate category include East Nusa Tenggara, Papua, West Papua, West Sulawesi, and South Sumatra. Papua Province recorded the highest average poverty rate in this group, which was 27,885, followed by West Papua with an average of 24,285, and South Sumatra which reached 23,103. West Sulawesi Province was slightly lower with an average poverty rate of 21,997, while East Nusa Tenggara had an average poverty rate of 20,864. These data show that the provinces in cluster 3 face significant poverty challenges, with variations in the level of social and economic disparities between regions. Poverty alleviation efforts in these provinces require special attention in order to reduce poverty rates effectively and sustainably.

nama_provinsi	rata_rata_kemiskinan	Cluster
MALUKU	18.86	Cluster 3 (Tinggi)
NUSA TENGGARA TIMUR	20.86	Cluster 3 (Tinggi)
PAPUA	27.88	Cluster 3 (Tinggi)
PAPUA BARAT	24.38	Cluster 3 (Tinggi)

REFERENCE

[1] Amarasinghe, U., Samad, M., & Anputhas, M. (2020). Spatial clustering of rural poverty and

food insecurity in Sri Lanka. *Food Policy*, *30*(5–6), 493–509. https://doi.org/10.1016/j.foodpol.2005.09.006

[2] Candra, Y., Goejantoro, R., & Dani, A. T. R. (2024). Pengelompokan Provinsi Berdasarkan Indikator Ekonomi, Pendidikan, Kesehatan, dan Kriminalitas di Indonesia Menggunakan Algoritma Centroid Linkage. *Euler : Jurnal Ilmiah Matematika, Sains Dan Teknologi*, *12*(1), 9–15. https://doi.org/10.37905/euler.v12i1.24887

[3] Chadli, F. E., Gretete, D., & Moumen, A. (2022). *Data Analysis within a Scientific Research Methodology*. 148–153. https://doi.org/10.5220/0010730000003101

[4] Chander, S., & Vijaya, P. (2021). Unsupervised learning methods for data clustering. In *Artificial Intelligence in Data Mining* (pp. 41–64). Elsevier. https://doi.org/10.1016/B978-0-12-820601-0.00002-1

[5] Chaudhry, M., Shafi, I., Mahnoor, M., Vargas, D. L. R., Thompson, E. B., & Ashraf, I. (2023a). A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective. In *Symmetry* (Vol. 15, Issue 9). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/sym15091679

[6] Chaudhry, M., Shafi, I., Mahnoor, M., Vargas, D. L. R., Thompson, E. B., & Ashraf, I. (2023b). A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective. In *Symmetry* (Vol. 15, Issue 9). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/sym15091679

[7] Erda, G., Gunawan, C., & Erda, Z. (2023). GROUPING OF POVERTY IN INDONESIA USING K-MEANS WITH SILHOUETTE COEFFICIENT. *Parameter: Journal of Statistics*, *3*(1), 1–6. https://doi.org/10.22487/27765660.2023.v3.i1.16435

[8] Foell, A., & Pitzer, K. A. (2020). Geographically Targeted Place-Based Community Development Interventions: A Systematic Review and Examination of Studies' Methodological Rigor. *Housing Policy Debate*, *30*(5), 741–765. https://doi.org/10.1080/10511482.2020.1741421

[9] Gonzalez, T. F. (2020). CLUSTERING TO MINIMIZE THE MAXIMUM INTERCLUSTER DISTANCE*. In *Theoretical Computer Science* (Vol. 38).

[10] Hasan, E., Rahman, Md. A., Shojib Talukder, MD., Utsho, M. F., Shakhan, Md., & Farid, D. Md. (2023). Data Segmentation with Improved K-Means Clustering Algorithm. *2023 26th International Conference on Computer and Information Technology (ICCIT)*, 1–5. https://doi.org/10.1109/ICCIT60459.2023.10441078

[11] He, Q., Chen, Z., Ji, K., Wang, L., Ma, K., Zhao, C., & Shi, Y. (2020). Cluster center initialization and outlier detection based on distance and density for the k-means algorithm. *Advances in Intelligent Systems and Computing*, *940*, 530–539. https://doi.org/10.1007/978-3-030-16657-1_49

[12] Hidayat, M. F., Wulandari, A., Hosea, H., Santoso Gunawan, A. A., & Permai, S. D. (2022). Analyzing Poverty Rate Using ANOVA in Indonesia. *2022 International Conference on Science and Technology (ICOSTECH)*, 1–4. https://doi.org/10.1109/ICOSTECH54296.2022.9829145

[13] Irani, J., Pise, N., & Phatak, M. (2020). Clustering Techniques and the Similarity Measures used in Clustering: A Survey. In *International Journal of Computer Applications* (Vol. 134, Issue 7).

[14] Islam, I., & Khan, H. (2020). Spatial patterns of inequality and poverty in indonesia. *Bulletin of Indonesian Economic Studies*, *22*(2), 80–102. https://doi.org/10.1080/00074918612331334834

[15] Kim, H., Kim, H. K., & Cho, S. (2020). Improving spherical k-means for document clustering: Fast initialization, sparse centroid projection, and efficient cluster labeling. *Expert Systems with Applications*, *150*. https://doi.org/10.1016/j.eswa.2020.113288

[16] Kliuchnyk, R. M. (2022). MAIN INTERPRETATIONS OF POVERTY IN ECONOMIC SCIENCE. *Academic Review*,

1(56), 14–23. https://doi.org/10.32342/2074-5354-2022-1-56-2

[17] Kurniasari, A., & Oktavilia, S. (2023). Andini Kurniasari & Shanty Oktavillia. *Economics Development Analysis Journal*, *12*(1). http://journal.unnes.ac.id/sju/index.php/edaj

[18] Li, H., Rajbahadur, G. K., & Bezemer, C.-P. (2024). Studying the Impact of TensorFlow and PyTorch Bindings on Machine Learning Software Quality. *ACM Transactions on Software Engineering and Methodology*. https://doi.org/10.1145/3678168

[19] Llerena-Izquierdo, J., Mendez-Reyes, J., Ayala-Carabajo, R., & Andrade-Martinez, C. (2024). Innovations in Introductory Programming Education: The Role of AI with Google Colab and Gemini. *Education Sciences*, *14*(12), 1330. https://doi.org/10.3390/educsci14121330

[20] Poerwanto, B. (2023). K-Means – Resilient Backpropagation Neural Network in Predicting Poverty Levels. *Jurnal Varian*, *6*(2), 157–166. https://doi.org/10.30812/varian.v6i2.2756

[21] Sano, A. V. D., & Nindito, H. (2020a). Application of K-Means Algorithm for Cluster Analysis on Poverty of Provinces in Indonesia. *ComTech: Computer, Mathematics and Engineering Applications*, *7*(2), 141. https://doi.org/10.21512/comtech.v7i2.2254

[22] Sano, A. V. D., & Nindito, H. (2020b). Application of K-Means Algorithm for Cluster Analysis on Poverty of Provinces in Indonesia. *ComTech: Computer, Mathematics and Engineering Applications*, *7*(2), 141. https://doi.org/10.21512/comtech.v7i2.2254

[23] Simon, G., Lee, J. A., & Verleysen, M. (2020). Unfolding preprocessing for meaningful time series clustering. *Neural Networks*, *19*(6–7), 877–888. https://doi.org/10.1016/j.neunet.2006.05.020

[24] Sumargo, B., & Haida, R. N. (2020). Linkages between Economic Growth, Poverty and Environmental Quality in Indonesia. *Jurnal Ekonomi Pembangunan: Kajian Masalah Ekonomi Dan Pembangunan*, *21*(1), 47–59. https://doi.org/10.23917/jep.v21i1.8262

[25] Timpe, L. C. (2023). An online <scp>G</scp> oogle <scp>C</scp> olab project for exploring the <scp>SARS CoV</scp> -2 genome and <scp>mRNA</scp> vaccines. *Biochemistry and Molecular Biology Education*, *51*(2), 209–211. https://doi.org/10.1002/bmb.21711

[26] Vrahatis, M. N., Boutsinas, B., Alevizos, P., & Pavlides, G. (2020). The new k-windows algorithm for improving the k-means clustering algorithm. *Journal of Complexity*, *18*(1), 375–391. https://doi.org/10.1006/jcom.2001.0633

[27] Winkler, W. E. (2020). Methods for evaluating and creating data quality. *Information Systems*, *29*(7), 531–550. https://doi.org/10.1016/j.is.2003.12.003

[28] Yu, S. S., Chu, S. W., Wang, C. M., Chan, Y. K., & Chang, T. C. (2020). Two improved k-means algorithms. *Applied Soft Computing Journal*, *68*, 747–755. https://doi.org/10.1016/j.asoc.2017.08.032