Comparing K-Means and DBSCAN Algorithms for Clustering Poverty Levels in Papua Islands

Katherine Vanessa
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480, Indonesia
katherine.vanessa@binus.ac.id

Irene Anindaputri Iswanto
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480, Indonesia
irene.iswanto@binus.ac.id

Katrina Wijaya
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480, Indonesia
katrina.wijaya@binus.ac.id

Muhammad Fadlan Hidayat
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480, Indonesia
muhammad.hidayat03@binus.ac.id

2024 9th International Conference on Information Technology and Digital Applications (ICITDA) | 979-8-3315-1073-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICITDA64560.2024.10810077

***Abstract*—**Poverty has become one of the main challenges in Indonesia, especially in the eastern regions. According to data from the Central Bureau of Statistics (BPS), Papua Island currently ranks as the poorest province. As of 2018, the poverty rate in Papua was 27.6%, and as of March 2023, it was 26.03%. The government has made various efforts to address the issue of poverty in Papua, one of which is providing financial assistance in the form of BLT (Direct Cash Assistance). However, to achieve sustainable development, more effective and targeted strategies are needed. The government's efforts to address poverty in Papua must be supported by accurate and efficient data analysis. This study involves clustering data on impoverished populations from 42 districts/cities in Papua and West Papua, obtained from the Central Bureau of Statistics (BPS) website in 2023. The study aims to compare the clustering results using the K-Means and DBSCAN algorithms. The performance of these algorithms is determined by the highest Silhouette Index (SI) value. After the data underwent a standardization process, the clustering results using the K-Means algorithm achieved the best SI value of 0.6359 with K=4. Meanwhile, the clustering results using the DBSCAN algorithm achieved the best SI value of 0.3678 with eps 1.8 and MinPts 8. The SI value of the K-Means algorithm is higher than that of the DBSCAN algorithm, indicating that clustering using the K-Means algorithm is superior to DBSCAN.

Keywords*—Poverty, K-Means Algorithm, DBSCAN Algorithm, Silhouette Index (SI)*

I. Introduction

The issue of poverty in Indonesia continues to increase and has not been successfully addressed to date [1]. Poverty is considered a challenge in addressing constraints in economic access, culture, politics, and community participation [2]. Its roots can arise from domestic as well as global situations. Poverty does not only refer to financial lack but also to the failure to meet the rights and proper treatment of every individual [3]. Therefore, steps to alleviate poverty need to be taken.

Poverty has become one of the main challenges in Indonesia, especially in the eastern part of the country. Based on data from the Central Statistics Agency (BPS), there are still at least ten provinces with poverty rates above 10%. Here are the top ten provinces with the highest percentage of poor population.

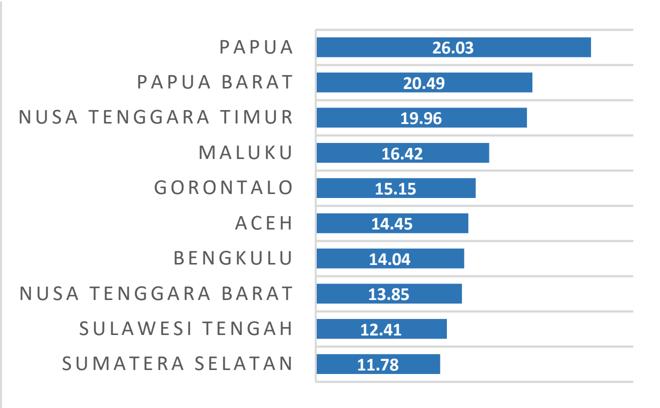

Figure 1. 10 Percentage of Largest Population Poverty Rates in Indonesia (Source: Badan Pusat Statistik, 2023)

Based on data from the Central Bureau of Statistics (BPS), Papua currently ranks as the poorest province in Indonesia. According to data presented in 2018, the poverty rate in Papua reached 27.6% [4], and as of March 2023, it was 26.03% based on the Papua Provincial Statistics Agency. These figures indicate that a significant portion of the population in Eastern Indonesia lives below the poverty line, necessitating special attention in poverty alleviation efforts. Selecting Papua as the focus of this research is important because it is one of the regions with the highest poverty rates in Indonesia [24]. Additionally, Papua's islands possess significant natural resource potential that has not yet been optimally utilized. Therefore, this study aims to apply clustering technology to improve the welfare of the people in Eastern Indonesia, particularly in the Papua Islands, through more effective data analysis.

The government has undertaken various efforts to address the issue of poverty in Papua. One of the government's measures to reduce poverty is by providing direct financial assistance, such as BLT, to those in need [5]. However, to achieve sustainable development, more effective and targeted strategies are required. One aspect that needs attention is the economic development in the Papua Islands. Sustainable economic development can have a positive impact on reducing poverty levels and improving the welfare of the population.

Government efforts to address poverty in Papua need to be supported by accurate and efficient data analysis. By using appropriate clustering methods, the government can identify groups of people vulnerable to poverty and design more targeted assistance programs. Clustering is the process of grouping several data points into two or more clusters until the data within the same group shows similarities based on the information contained in the data points [6]. Several algorithms are used in clustering, such as K-Means and DBSCAN. The goal is to reduce the sensitivity of the partitioning results to extreme values in the dataset by using the medoid as the representative point of each cluster, rather than based on the average of the observed data points [7].

Previous research has used clustering methods such as K-Means, and DBSCAN in economic development. For example, a study evaluating the use of K-Means and K-Medoids in clustering poverty data in Indonesia showed that K-Means was more effective [5]. The use of DBSCAN has also proven successful in producing two clusters with significant characteristic differences in the context of rice farming land in Karawang Regency [8]. These research findings provide additional insights into the application of clustering methods in economic development.

One of the main challenges in clustering poverty data is its irregular distribution and the presence of noise and outliers, which complicate the process of forming meaningful clusters. This issue is compounded by the fact that different clustering methods, such as K-Means and DBSCAN, may perform differently depending on the nature of the data, especially when standardization is applied.

In this study, we will compare the effectiveness of clustering methods, specifically K-Means and DBSCAN, in grouping poverty data in the Papua Islands. K-Means is chosen because it has proven effective and is commonly used in data analysis. Meanwhile, DBSCAN is selected for its ability to handle data with noise and outliers, which are often challenges in poverty data analysis [8]. The comparison between K-Means and DBSCAN is necessary due to several factors. Poverty data in Papua may be irregularly distributed which challenges K-Means' assumption of spherical clusters. With standardized data, the comparison between K-Means and DBSCAN remains necessary, but the focus shifts slightly. While standardization helps K-Means by scaling features, it still assumes spherical clusters [20], whereas DBSCAN handles noise and identifies clusters of any shape [19]. K-Means requires predefined clusters [20] and is computationally efficient [21], while DBSCAN determines clusters based on density [19]. Thus, even with standardized data, comparing these methods helps identify the most effective approach for capturing complex patterns and informs better poverty alleviation strategies.

Through this research, it is hoped that a better understanding will be obtained regarding the more effective application of clustering methods among the two in the context of poverty in Papua, to achieve optimal results, as well as contributing to the development of more effective development strategies to improve the welfare of the people in Eastern Indonesia, particularly in Papua.

II. Literature Review

A. Previous Work

According to previous research conducted by [13], applying the K-Means and K-Medoids methods to the topic of poverty in Sumatra Island, the K-Means algorithm was considered slightly superior to the K-Medoids algorithm. This was because the K-Means algorithm produced a smaller Davies-Bouldin Index (DBI) score compared to the K-Medoids algorithm. For further research by [4] on clustering poverty data in Indonesia using the K-Means and K-Medoids algorithms, similar results were obtained, with K-Means being deemed superior to K-Medoids. Both studies could consider using alternative validation methods, such as the Silhouette Index, and could be expanded by applying poverty data from other regions.

The next tested algorithm is the DBSCAN algorithm. DBSCAN clusters objects based on the density of objects in an area [10]. Research on the DBSCAN method with K-Means has not been widely conducted yet. According to previous research that used the DBSCAN and K-Means algorithms, such as in the case of COVID-19 conducted by [11], the results showed that the K-Means algorithm with a Silhouette Index value of 0.6902 and $k = 8$ was superior to the DBSCAN algorithm. From the research, both algorithms can be further developed for use with data that has a complex structure, such as spatial data or data with significant variations in density.

Meanwhile, a study comparing K-Means and DBSCAN for Text Clustering of Product Reviews showed that DBSCAN achieved an accuracy of 99.80%, slightly higher than K-Means, which had 99.50% [23]. The main factor behind this difference is the variation in parameter settings, resulting in different numbers of clusters. The results could be improved with better data preprocessing, as the research indicated that DBSCAN identified a lot of noise in the data, and some clusters did not produce optimal results.

Therefore, this study will further discuss the best algorithm for clustering with the topic of poverty levels in Papua. In several previous studies discussing similar topics, K-Means was considered the better algorithm to be implemented in such topics. However, there are not many studies on similar topics with DBSCAN algorithms. Hence, this research is conducted to compare which algorithm yields more accurate results among the two algorithms for the topic of poverty research in a specific region.

B. Clustering

Clustering is a method in data mining aimed at grouping several objects into specific clusters. With clustering, objects with similar characteristics are placed into the same cluster, while objects with different characteristics are placed into different clusters [22]. Unlike classification, clustering does not require labels, hence it is also known as unsupervised learning [15]. Clustering analysis is used to gather information as closely as possible, making it easier to make conclusions without reducing the original information.

There are two types of clustering methods: hierarchical data clustering and non-hierarchical data clustering [9]. The clear difference between hierarchical and non-hierarchical methods is that hierarchical methods perform grouping gradually, while non-hierarchical methods perform clustering on sample space using partitioning methods [14].

C. K-Means Clustering

K-Means is a non-hierarchical clustering method, the simplest and most commonly used clustering method. K-means is used to partition similar objects and map them to the nearest centroid to form a cluster. Meanwhile, objects that are far from the centroid will be grouped into different clusters [13].

The K-Means algorithm has several advantages for clustering data [16][17]:

1. Conceptually simple and fast to execute, especially on large datasets.
2. Works well when clusters are spherical or well-separated because it minimizes the distance to the centroid.
3. It scales well to large datasets and efficient, suitable for real-time applications.
4. Quickly converges with predefined clusters and initialization.

The following are the steps of the K-Means Clustering algorithm [15] outlined as follows:

1. Determine the data to be objects and the number of clusters, X_{ij} (i =1,2,....n; j = 1,2,....m) in the equation where n represents the number of objects while m represents the number of data variables..
2. In the first iteration, randomly determine the centers (centroids), C_{kj} (k = 1,.....k;j = 1,.....m).
   Determine the distance between the object and the centroid using the Euclidean formula:

$$d_{ik} = \sqrt{\sum_{j=1}^{m} \left(X_{ij} - C_{kj}\right)^2} \quad (1)$$

3. Objects with a minimum distance to the cluster center are grouped.
4. Calculate the cluster centers with the following formula, if when the next iteration with the new cluster centers still results in data changes, then iterate again. The following is the formula for the new centroid:

$$C_{kj} = \frac{\sum_{i=1}^{p} X_{ij}}{p} \quad (2)$$

5. If there are no data changes in other clusters, then the iteration process can be stopped.

D. DBSCAN Clustering

Density-Based Spatial Clustering of Application with Noise or DBSCAN is a clustering algorithm that groups objects based on data density. DBSCAN is an algorithm with a non-hierarchical clustering method which is considered a popular algorithm as an alternative to the K-Means algorithm [12]. The advantage of the DBSCAN algorithm compared to K-Means is that there is no limitation on the number of clusters set during initialization. Instead, the number of clusters is determined based on the data density area.

The DBSCAN algorithm has several advantages for clustering data [17][18]:

1. Effective in handling outliers and noise in the data.
2. Find clusters of arbitrary shapes and sizes, makes DBSCAN more versatile in handling complex data structures.
3. Does not require the number of clusters to be specified in advance.
4. Operates on a density-based approach to identify clusters.

The following are the steps of the DBSCAN algorithm [10] outlined as follows:

1. Determine the initialization value of the MinPts and Eps parameters.
2. Randomly determine the value of p (or starting point).
3. Calculate Eps or the distance of all points within the density range to the starting point or p using the Euclidean formula:

$$d_{ij} = \sqrt{\sum_{a}^{p} \left(X_{ia} - X_{ja}\right)^2} \quad (3)$$

4. The point p is considered as a core point and a cluster is formed when the number of points that satisfy Eps is greater than MinPts.
5. Itterate steps 3-4 iteratively until the entire process is performed on all available points. If p is a border point and there are no points within the reachable density, then the process continues to the next point.

III. METHODOLOGY

The research methodology utilizes a systematic approach to conduct the experiment, with various stages outlined in Figure 2. These stages can be described in the following order:

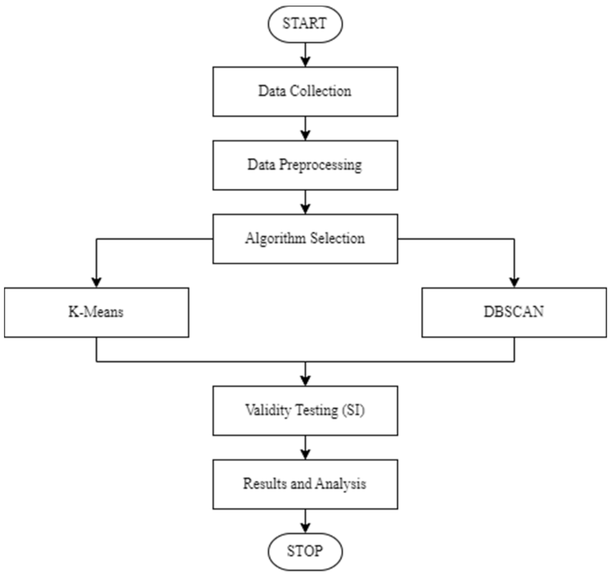

Figure 2. Research Methodology

A. Data Collection

In this study, the data used is the data on impoverished populations obtained from the Central Bureau of Statistics (BPS) through the BPS.go.id website in 2023. The variables used in this research are the data of districts/cities in the Provinces of Papua and West Papua in 2023. This study considers two main indicators, namely monetary and non-monetary indicators, as outlined by Amartya Sen in his work "Development as Freedom" (1999). The variables included in each indicator are as follows:

a. Monetary Indicator

According to Amartya Sen in his book "Development as Freedom" (1999), the monetary indicator is often used in traditional poverty approaches that focus on income or expenditure as the primary measures of poverty.

TABLE I. MONETARY INDICATOR

Variable	*Description*	*Unit*	*Scale*
X1	Percentage of Poor Population	Ratio	Percentage
X2	Per Capita Expenditure	Ratio	Percentage

b. Non-Monetary Indicator

Amartya Sen also emphasizes the importance of capabilities and functions, which include non-monetary aspects. The Multidimensional Poverty Index (MPI) developed by the UNDP and OPHI also incorporates non-monetary indicators. Non-monetary indicators encompass aspects such as health, education, and living standards that are indirectly related to income but significantly affect quality of life.

TABLE II. NON-MONETARY INDICATOR

Variable	Description	Unit	Scale
X3	School Participation Rate (ages 13-15)	Ratio	Percentage
X4	Housing Facilities with Adequate Water Supply	Ratio	Percentage
X5	Housing Facilities with Private Toilets	Ratio	Percentage

The districts/cities in Papua and West Papua that will be grouped based on economic development indicators consist of 42 regions with a total of 39 districts and 3 cities. The data structure used in this study is as follows.

TABLE III. DATA STRUCTURE

No.	Province	X1	X2	X3	X4	X5
1	Fakfak	21,38	49,46	100	97,81	66,69
2	Kaimana	14,57	48,91	92,95	77,73	64,23
3	Teluk Wondama	28,9	53,85	98,99	38,75	33,67
4	Teluk Bintuni	28,24	52,43	97,05	81,83	68,29
5	Manokwari	18,73	44,46	100	72,6	96,19
…	…	…	…	…	…	…
42	Kota Jayapura	10,5	42,16	96,97	90,6	88,58

B. Data Preprocessing

After the data is collected, the first step is to perform standardization to ensure that all variable values fall within the same range. This standardization aims to improve the effectiveness of the clustering process. The standardized data results are displayed in the following table.

TABLE IV. STANDARDIZED DATA

No.	Province	X1	X2	X3	X4	X5
1	Fakfak	-0,5839	-0,8984	0,9071	1,2494	0,1644
2	Kaimana	-1,3706	-0,9602	0,5287	0,5120	0,0728
3	Teluk Wondama	0,2847	-0,4050	0,8529	-0,9192	-1,0648
4	Teluk Bintuni	0,2085	-0,5646	0,7488	0,6626	0,2239
5	Manokwari	-0,8900	-1,4603	0,9071	0,3237	1,2625
…	…	…	…	…	…	…
42	Kota Jayapura	-1,8407	-1,7188	0,7445	0,9846	0,9792

C. K-Means Clustering

This method clusters data so that data with similar characteristics are grouped together, while data with different characteristics are placed in different clusters. The steps involved in K-Means Clustering analysis are as follows:

1. Reading the csv file.
2. Calculating the average for each region.
3. Performing the clustering algorithm with iterations until a tolerance of 1e-4.
4. Displaying the plot of clustering results.

D. DBSCAN Clustering

DBSCAN is a clustering algorithm that groups objects based on data density, handling noise and outliers. The steps involved in DBSCAN Clustering analysis are as follows:

1. Reading the csv file
2. Iterating to find the parameter of epsilon and minimum points values that yield the best SI value.
   - Epsilon (ε): The maximum distance between two samples for one to be considered as in the neighborhood of the other.
   - MinPts: The minimum number of samples in a neighborhood for a point to be considered as a core point.
3. Performing the DBSCAN algorithm with the obtained epsilon and minimum points values.
4. Creating a plot of clustering results.

E. Silhouette Index (SI)

The validity of the data from these two algorithms was tested using the Silhouette Index (SI), an internal validation method that evaluates cluster placement by comparing the average distance of objects within the same cluster to those in different clusters according to X. Wang and Y. Xu (2019) [25]. The algorithm with the highest Silhouette Index value will be determined as the best algorithm.

IV. RESULTS AND ANALYSIS

A. Cluster Testing with the K-Means Algorithm

After standardizing the data, clustering was performed using the K-Means algorithm. The K-Means clustering was iterated to achieve a tolerance result of 0.0001, with trials of 2, 3, 4, and 5 clusters. The clustering results were then validated using the Silhouette Index (SI) to find the best clustering outcome. The graph below shows the clustering test results using the K-Means algorithm.

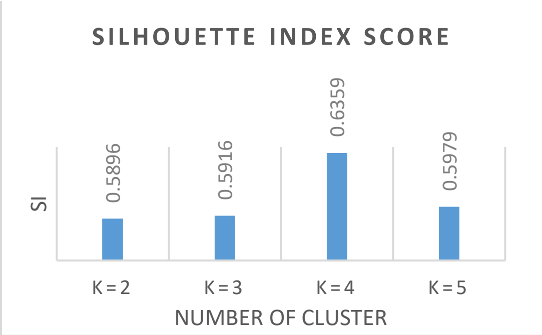

Figure 3. Graph of The Clustering Result Using SI for K-Means Algorithm

Based on the table, the highest SI value from clustering using the K-Means algorithm is 0.6359, which is achieved with K = 4. The results indicate that 15 regions are grouped in Cluster 1, 4 regions in Cluster 2, 17 regions in Cluster 3, and 6 regions in Cluster 4.

While the elbow method suggested that the optimal number of clusters is K = 3, this recommendation was not followed. The reason is that K = 4 provides a higher SI value compared to K = 3, indicating better clustering performance with four clusters. The results from the elbow method are as follows.

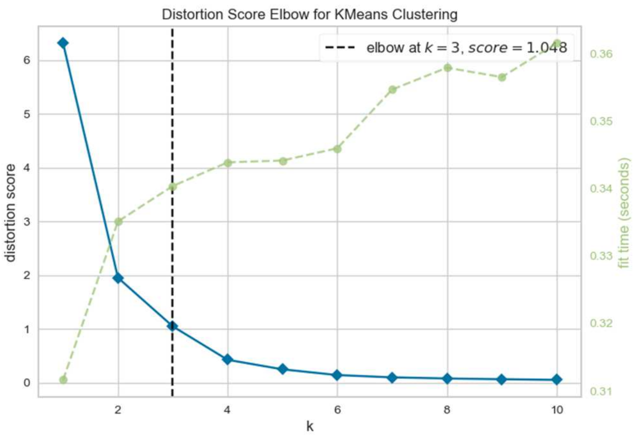

Figure 4. Graph of The Clustering Result Using Elbow Method for K-Means Algorithm

Based on the table, the best SI value from clustering using the K-Means algorithm is 0.6359, achieved with K = 4. The clustering results show that 15 regions belong to Cluster 1, 4 regions belong to Cluster 2, 17 regions belong to Cluster 3, and 6 regions belong to Cluster 4. The plot results can be seen next to this.

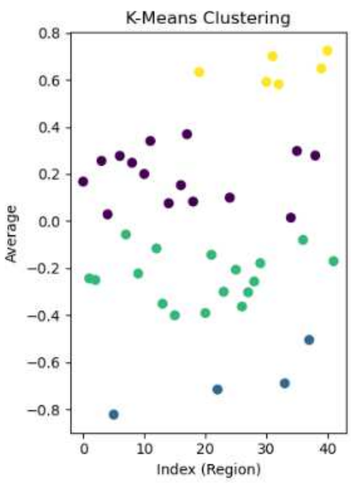

Figure 5. K-Means Algorithm Clustering Plot Results

B. Cluster Testing with the DBSCAN Algorithm

In addition to using the K-Means algorithm, clustering was also performed using the DBSCAN algorithm. Cluster iteration was conducted to find the best Eps and MinPts values. The MinPts values used in this iteration ranged from 1 to 10, and the Eps values ranged from 1.1 to 2. Afterward, the Silhouette Index validation test was conducted to evaluate the DBSCAN clustering results. The graph below shows the DBSCAN clustering test results using the best MinPts value of 8, and Eps ranging from 1.1 to 2.

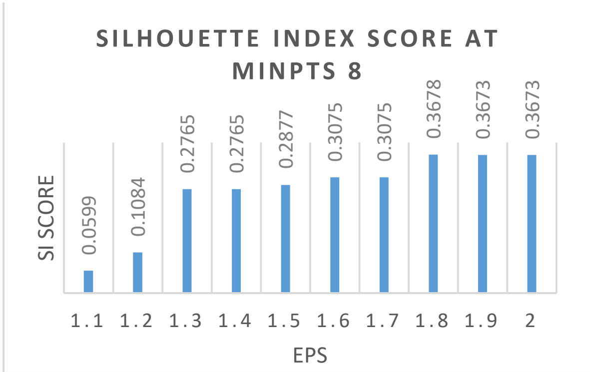

Figure 6. Graph of The Clustering Result Using SI for DBSCAN Algorithm

Based on the table, the best SI value from clustering using the DBSCAN algorithm is 0.3678, achieved with Eps of 1.8 and MinPts of 8. The clustering results show that 31 regions belong to Cluster 1, and 11 regions are classified as noise or outliers. Below is the clustering plot result.

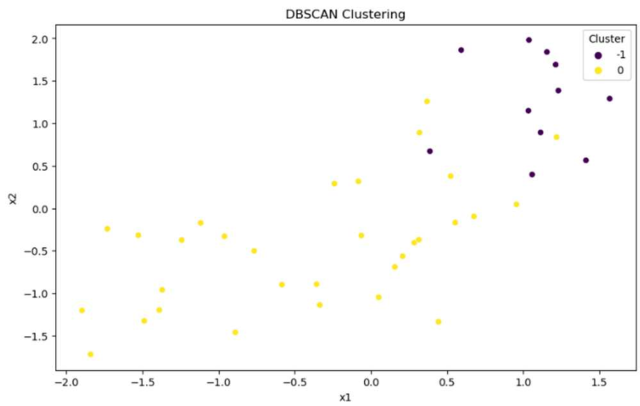

Figure 7. DBSCAN Algorithm Clustering Plot Results

V. CONCLUSIONS

During the clustering process, we faced challenges such as features in the multivariable dataset having different scales, which could result in larger scale features dominating distance calculations if the data is not standardized. Additionally, clustering results can be difficult to analyze if the cluster patterns are unclear. To address these issues, we standardized the data to ensure all features contribute equally, applied validation using the Silhouette Score, and used visualization to facilitate analysis. We also fine-tuned key parameters, such as the number of clusters in K-Means and the epsilon and MinPts values in DBSCAN, to achieve the best results.

Based on the validation tests of clustering results on the poverty rate data of districts/cities in Papua Island using the K-Means and DBSCAN algorithms, this study found that clustering using the K-Means algorithm is superior to DBSCAN. After the data standardization process, the clustering results from the K-Means algorithm achieved the best SI value of 0.6359 with K=4. In contrast, the DBSCAN algorithm achieved its best SI value of 0.3678 with eps 1.8 and MinPts 8.

In this case, K-Means outperforms DBSCAN. K-Means is more effective for data with structured clustering patterns, while DBSCAN is better suited for data with clear density. K-Means, with K = 4, yields the best results due to its ability to determine the optimal number of clusters. Conversely, DBSCAN relies on epsilon and min_samples parameters, which, if not set properly, can lead to less satisfactory clusters or excessive noise. Overall, the differences in clustering results may stem from variations in data structure, chosen parameters, and how each algorithm handles noise and cluster shapes. This indicates that the K-Means algorithm is more suitable for clustering districts/cities in Papua Island based on poverty indicators compared to the DBSCAN algorithm, as the SI value obtained is higher than the SI value from the DBSCAN algorithm.

Through the application of clustering, the government can identify poverty patterns across various regions, enabling more targeted interventions in areas with high poverty clusters (such as cluster 4). Additionally, clustering facilitates more precise resource allocation, tailored to the urgent needs of each cluster, such as infrastructure, education programs, or social assistance. This approach allows the government to design more effective policies to reduce economic disparities between regions. The patterns from this study can serve as a reference for clustering modelling to support the development of more effective strategies to improve the welfare of communities in Eastern Indonesia, particularly in Papua.

REFERENCES

[1] D. N. Alfiansyah, V. R. S. Nastiti, and N. Hayatin, “Penerapan Metode K-Means pada Data Penduduk Miskin Per Kecamatan Kabupaten Blitar,” J. Repos., vol. 4, no. 1, pp. 49–58, 2022, doi: 10.22219/repositor.v4i1.1416.

[2] I. Nasution, A. P. Windarto, and M. Fauzan, “Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi,” Build. Informatics, Technol. Sci., vol. 2, no. 2, pp. 76–83, 2020, doi: 10.47065/bits.v2i2.492.

[3] G. Dwilestari, Mulyawan, Martanto, And I. Ali, “Analisis Clustering Menggunakan K-Medoid Pada Data Penduduk Miskin Indonesia,” Jursima J. Sist. Inf. Dan Manaj., vol. 9, vo. 3, pp. 282290, 2021.

[4] N. T. Luchia, H. Handayani, F. S. Hamdi, D. Erlangga, and S. F. Octavia, “Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia,” MALCOM Indones. J. Mach. Learn. Comp. Sci., vol. 2, no. 2, pp. 35–41, 2022, doi: 10.57152/malcom.v2i2.422.

[5] F. Febriyana, "Analisis Klaster K-Means dan K-Median Pada Data Indikator Kemiskinan," Skripsi, Fakultas Ilmu Komputer, Universitas XYZ, 2011.

[6] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan,” J. Nas. Teknol. dan Sist. Inf., vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.

[7] I. Kamila, U. Khairunnisa, and M. Mustakim, “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau,” J. Ilm. Rekayasa dan Manaj. Sist. Inf., vol. 5, no. 1, p. 119, 2019, doi: 10.

[8] B. N. Sari and A. Primajaya, “Penerapan Clustering Dbscan Untuk Pertanian Padi Di Kabupaten Karawang,” J. Inform. dan Komput., vol. 4, no. 1, pp. 28–34, 2019, [Online]. Available: www.mapcoordinates.net/en.24014/rmsi.v5i1.7381.

[9] Yudi Agusta, “K-Means – Penerapan, Permasalahan dan Metode Terkait,” J. Sist. dan Inform., vol. 3, no. Februari, pp. 47–60, 2007.

[10] Ni, K. Gede, and I Made Sukarsa, “Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan,” Lontar Komputer, pp. 185–185, Dec. 2015, doi: https://doi.org/10.24843/lkjiti.2015.v06.i03.p05.

[11] R. Adha, N. Nurhaliza, U. Sholeha, and M. Mustakim, “Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia,” SITEKIN J. Sains, Teknol. dan Ind., vol. 18, no. 2, pp. 206–211, 2021, [Online]. Available: https://ejournal.uin-suska.ac.id/index.php/sitekin/article/view/12469

[12] A. A. P. Sari, “Implementasi Metode Improved K-Means,” pp. 1–69, 2020.

[13] Mardhiatul Azmi, Atus Amadi Putra, Dodi Vionanda, and Admi Salma, “Comparison of the Performance of the K-Means and K-Medoids Algorithms in Grouping Regencies/Cities in Sumatera Based on Poverty Indicators,” UNP J. Stat. Data Sci., vol. 1, no. 2, pp. 59–66, 2023, doi: 10.24036/ujsds/vol1-iss2/25.

[14] Cendiana Aprilia Haryono, “Penerapan Metode C-Means dan Fuzzy C-Means pada Pengelompokkan Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Indikator Pembangunan Ekonomi,” pp. 1–113, 2016.

[15] N. Butarbutar, A. P. Windarto, D. Hartama, and S. Solikhun, “Komparasi Kinerja Algoritma Fuzzy C-Means Dan K-Means Dalam Pengelompokan Data Siswa Berdasarkan Prestasi Nilai Akademik Siswa,” Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform., vol. 1, no. 1, p. 46, 2017, doi: 10.30645/jurasik.v1i1.8.

[16] B. Chong, “K-Means Clustering Algorithm: A Brief Review,” Academic Journal of Computing & Information Science, vol. 4, no. 5, Jan. 2021, doi: 10.25236/ajcis.2021.040506.

[17] N. P. Sutramiani, I. M. T. Arthana, P. F. Lampung, S. Aurelia, M. Fauzi, and I. W. A. S. Darma, “The Performance Comparison of DBSCAN and K-Means Clustering for MSMEs Grouping based on Asset Value and Turnover,” Journal of Information Systems Engineering and Business Intelligence, vol. 10, no. 1, pp. 13–24, Feb. 2024, doi: 10.20473/jisebi.10.1.13-24.

[18] N. Nurhaliza and M. Mustakim, “Pengelompokan Data Kasus Covid-19 di Dunia Menggunakan Algoritma DBSCAN,” Indonesian Journal of Informatic Research and Software Engineering (IJIRSE), vol. 1, no. 1, pp. 1–8, Apr. 2021, doi: 10.57152/ijirse.v1i1.42.

[19] A. Ram, S. Jalal, A. S. Jalal, and M. Kumar, “A Density Based Algorithm for Discovering Density Varied Clusters in Large Spatial Databases,” International Journal of Computer Applications, vol. 3, no. 6, pp. 1–4, Jun. 2010, doi: 10.5120/739-1038.

[20] P. J. Rousseeuw, “Silhouettes: A Graphical Aid to The Interpretation and Validation of Cluster Analysis,” Journal of Computational and Applied Mathematics, vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.

[21] J. B. MacQueen, “Some Methods for Classification and Analysis of Multivariate Observations,” Project Euclid, Jan. 01, 1967. http://projecteuclid.org/euclid.bsmsp/1200512992

[22] K. Handoko and L. S. Lesmana, “Data Mining Pada Jumlah Penumpang Menggunakan Metode Clustering,” Prosiding Seminar Nasional Ilmu Sosial Dan Teknologi (SNISTEK), vol. 1, pp. 97–102, Oct. 17, 2018. https://ejournal.upbatam.ac.id/index.php/prosiding/article/view/751

[23] F. Andriyani and Y. Puspitarani, “Performance comparison of K-Means and DBScan algorithms for text clustering product reviews,” SinkrOn, vol. 7, no. 3, pp. 944–949, Jul. 2022, doi: 10.33395/sinkron.v7i3.11569.

[24] G. Haumahu and M. Y. Matdoan, “Algoritma K-Medoids Clustering untuk Mengelompokkan Tingkat Kemiskinan Pada Kabupaten dan Kota Di Kepulauan Maluku dan Papua,” VARIANCE Journal of Statistics and Its Applications, vol. 4, no. 2, pp. 81–88, Mar. 2023, doi: 10.30598/variancevol4iss2page81-87.

[25] X. Wang and Y. Xu, “An Improved Index for Clustering Validation Based On Silhouette index and Calinski-Harabasz index,” IOP Conference Series Materials Science and Engineering, vol. 569, no. 5, p. 052024, Jul. 2019, doi: 10.1088/1757-899x/569/5/052024.