Jurnal Sistem Komputer dan Informatika (JSON)
Volume 6, Nomor 4, Juni 2025
e-ISSN 2685-998X
DOI 10.30865/json.v6i4

Implementation of DBSCAN Algorithm for Grouping Poverty Levels in Central Java Province

Maulida Aristia Tsani, Amiq Fahmi*

Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1]112202106612@mhs.dinus.ac.id, [2*]amiq.fahmi@dsn.dinus.acc.id
Correspondence Author Email: amiq.fahmi@dsn.dinus.ac.id*
Submitted: **24/04/2025**; Accepted: **26/05/2025**; Published: **30/06/2025**

Abstrak− Kemiskinan merupakan masalah kompleks yang menghambat pembangunan sosial ekonomi di Indonesia, khususnya di Provinsi Jawa Tengah, yang menghadapi tantangan serius dengan tingkat kemiskinan mencapai 10,77% pada tahun 2023. Kondisi ini mencerminkan ketimpangan distribusi yang signifikan antarwilayah. Penelitian ini bertujuan untuk mengeksplorasi persebaran kemiskinan secara spasial di 35 kabupaten/kota di Jawa Tengah melalui segmentasi wilayah yang didasarkan pada tingkat kemiskinan. Data yang diolah berasal dari sistem Badan Pusat Statistik Provinsi Jawa Tengah tahun 2023. Variabel yang digunakan dalam penelitian ini yaitu Kabupaten/Kota, Upah Minimum Kabupaten (UMK), dan tingkat pengangguran. Dalam mengidentifikasi pola spasial kemiskinan memanfaatkan algoritma *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN). Hasil tinjauan mengungkapkan empat klaster, yaitu klaster 0 (kemiskinan sedang) terdapat 28 kabupaten/kota, klaster 1 (kemiskinan tinggi) terdapat 2 kabupaten/kota, klaster 2 (kemiskinan sangat tinggi) terdapat 1 kabupaten/kota, dan klaster 3 (kemiskinan rendah) terdapat 4 kabupaten/kota. Evaluasi yang digunakan yaitu *Silhouette Score* dan *Davies-Bouldin Index* yang masing-masing menghasilkan 0,447 dan 0,441, menunjukkan keterpisahan yang cukup baik antar klaster, serta memberikan gambaran yang lebih terstruktur mengenai distribusi kemiskinan di Jawa Tengah. Hasil penelitian ini mampu mengelompokkan tingkat kemiskinan berdasarkan Upah Minimum Kabupaten (UMK) dan Tingkat pengangguran dengan menggunakan algoritma DBSCAN.

Kata Kunci: Kemiskinan; Provinsi Jawa Tengah; Pengelompokan DBSCAN; Skor Silhouette; Indeks Davies-Bouldin

Abstract− Poverty is a complex problem that hinders socio-economic development in Indonesia, especially in Central Java Province, which is an anomaly in a serious problem with a poverty rate of 10.77% in 2023. This condition reflects significant inequality in distribution between regions. This study intends to explore the spatial distribution of poverty in 35 districts/cities in Central Java through regional segmentation according to poverty levels. The processed data were obtained from the official system of the Central Java Provincial Statistics Agency in 2023. The variables used in this study are Regency/City, Regency Minimum Wage (UMK), and Poverty Level. In identifying spatial poverty patterns using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm. The review's outcome revealed four clusters, specifically cluster 0 (moderate poverty), which was constituted by 28 districts/cities; cluster 1 (high poverty), which was constituted by 2 districts/cities; cluster 2 (very high poverty), which was constituted by 1 district/city; and cluster 3 (low poverty), which consisted of 4 districts/cities. The evaluations used were the Silhouette Score and Davies-Bouldin Index, which produced 0.447 and 0.441, respectively, indicating a reasonably good separation between clusters and providing a more structured picture of the distribution of poverty in Central Java. The results of this study were able to group poverty levels based on the District Minimum Wage (UMK) and Poverty Level using the DBSCAN algorithm.

Keywords: Poverty; Central Java Province; DBSCAN Clustering; Silhouette Score; Davies-Bouldin Index

1. INTRODUCTION

An individual or group is in poverty if they cannot meet fundamental conditions such as adequate health care, food, clothing, shelter, and education due to financial constraints [1]. Poverty is a significant challenge in socio-economic and development aspects, especially in developing countries such as Indonesia [2]. Prolonged poverty can hinder national progress and worsen social inequality.

Poverty has a substantial impact on social stability and regional equality, in addition to the economy. Increasing social injustice, limited access to education and health services, and unequal distribution of economic resources in society all of which can lead to social instability and violence [3], [4]. Addressing this requires a targeted poverty alleviation strategy encompassing economic empowerment, enhanced access to education, and improved health services [5]. These strategies aim to produce a more equitable society by addressing the root causes of poverty. Fostering social cohesion and promoting sustainable development can be achieved through a focus on empowerment and access.

Data mining is a strategy that combines statistical methods, mathematics, artificial intelligence, and machine learning to identify patterns or valuable information from large datasets [6]. One of the methods typically used in data mining is the clustering method, which divides data into various groupings depending on the similarity of features between data objects [7]. This approach can identify hidden patterns in data to understand the characteristics of the clusters formed [8]. The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm is utilized in this study for its ability to cluster data based on density effectively [9]. This algorithm is deemed relevant due to its capability to manage unevenly distributed data, reflecting the characteristics of poverty across regions in Central Java.

A study by [10] employed clustering techniques using the K-Means algorithm, focusing on poverty data from East Java Province obtained from the East Java Central Statistics Agency (BPS) between 2021 and 2023.

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 6, Nomor 4, Juni 2025
e-ISSN 2685-998X
DOI 10.30865/json.v6i4

The study utilized variables such as the inadequate population and unemployment rate, inducing 2 clusters with evaluation metrics including a Silhouette Score of 0.550 and a Davies-Bouldin Index of 0.600. However, this research was limited to East Java and did not incorporate the DBSCAN algorithm. Similarly, [11] applied the K-Means algorithm to Central Java Province's poverty data from the BPS in 2022, using attributes such as the number of poor individuals, population growth rate, and open unemployment rate. This study grouped districts and cities into 3 clusters but did not employ advanced metrics such as the Silhouette Score and Davies-Bouldin Index or consider the DBSCAN algorithm.

In contrast, [12] utilized the K-Medoids algorithm on a dataset encompassing poverty data from 34 Indonesian provinces sourced from the BPS between 2015 and 2022. The dataset included 10 variables across 272 rows, producing 3 clusters, specifically very poor (50%), poor (45%), and vulnerable poor (5%). While evaluated using a Silhouette Score of 0.4, this research did not apply the DBSCAN algorithm and was restricted to national-level clustering. Meanwhile, [13] clustered earthquake points on Sulawesi Island using the ST-DBSCAN algorithm. Their dataset, sourced from the Meteorology, Climatology, and Geophysics Agency (BMKG) database 2019, comprised variables such as longitude, latitude, and earthquake date. This study identified 60 clusters and 216 noise points, evaluated with a Spatial Silhouette Coefficient of 0.0569 and a Temporal Silhouette Coefficient of -0.5723. However, the study's focus remained solely on earthquake grouping.

Focused DBSCAN applications include the research by [14], which analyzed disaster-prone regions on Sumatra Island using natural disaster data from the National Disaster Management Agency (BNPB) between 2015 and 2021. Utilizing parameters Eps = 0.18 and MinPts = 4, this study identified 2 clusters and 44 noise points, with evaluations relying on a Silhouette Score of 0.46. However, this research lacked metric diversity and emphasized disaster-prone area grouping. Similarly, [15] applied DBSCAN to poverty data from West Kalimantan Province sourced from the BPS in 2021. Using variables such as the Human Development Index (HDI), literacy rates, and open unemployment rates, this study produced two clusters and five noise points with a Silhouette Coefficient of 0.672. While effective, the research was limited to West Kalimantan and lacked comparative evaluation metrics, illustrating the need for broader exploration and methodological advancements.

This study aims to utilize the DBSCAN algorithm to sort districts/cities in Central Java according to poverty rates and evaluate the usefulness of the DBSCAN method in sorting areas depending on poverty data density. DBSCAN was selected for its proficiency in managing unevenly distributed data, making it particularly suitable for uncovering intricate poverty patterns comprehensively [16]. The DBSCAN algorithm to provide a more in-depth analysis of poverty patterns by identifying and revealing spatial variations between regions. The findings of this research provide the ideal cluster count based on the poverty levels of districts/cities in Central Java Province. Thus, this clustering offers insight for the government in formulating more effective and targeted policies. In addition, involving the DBSCAN algorithm increases the accuracy and effectiveness of poverty grouping analysis in Central Java Province. This research is expected to produce a grouping of regions based on social organizations to develop more effective and targeted poverty alleviation strategies.

2. RESEARCH METHODOLOGY

The research method in this study consists of four main stages: dataset collection, data preprocessing, clustering using the DBSCAN algorithm (Density Clustering Spatial Based on Application with Noise), and evaluation. All of these stages are presented in Figure 1.

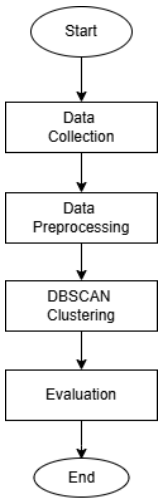


Figure 1. Stages of Research Methods

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 6, Nomor 4, Juni 2025
e-ISSN 2685-998X
DOI 10.30865/json.v6i4

2.1 Data Collection

The dataset in this study was taken from the Central Statistics Agency (BPS) website of Central Java Province. The dataset contains information on poverty data in Central Java Province in 2023 with three main attributes: Regency/City, District Minimum Wage (UMK), and Unemployment Rate. These three attributes provide comprehensive information on the factors influencing Central Java Province's poverty levels. An example of a sample poverty dataset is presented in Table 1.

Table 1. Poverty Dataset Sample

District/City	District Minimum Wage	Unemployment Rate
Banjarnegara	2038005	6.26
Banyumas	2118125	6.35
Batang	2284627	6.06
…	...	...
…	…	…
Wonogiri	1968448	1.92
Wonosobo	2076209	4.95

2.2 Data Preprocessing

The data preprocessing phase is designed to validate the suitability and readiness of the collected data to be analyzed in the next phase. Irrelevant entries, including missing values, incomplete records, or duplicates, may lead to analytical errors. To resolve these problems, the following data preprocessing steps are implemented.

2.2.1 Data Cleaning

Data cleansing is crucial in data preparation to improve quality by correcting inaccuracies, incompleteness, and inconsistencies. This step is conducted to validate the quality of the data utilized in the analysis has a high level of accuracy. Data cleansing involves assessing the quality by changing, removing, or correcting irrelevant, inconsistent, incomplete, or invalid information [17]. Some basic techniques that are applied include filling in empty values using the median or mode according to the type and distribution of the data, deleting duplicate data that can affect the analysis results, and handling outliers by deleting or replacing their values to fit within reasonable limits.

2.2.2 Data Transformation

Data transformation is changing data into a structure suitable for analysis to support decision-making [18]. This process aims to improve data quality and consistency so that the analysis model makes it easier to understand and use. Regular techniques used in this stage include normalization, standardization, and encoding, depending on the analysis's needs.

2.3 DBSCAN Clustering

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based clustering algorithm that detects outliers and handles noisy data. This algorithm does not require an initial identification of the optimal cluster count and is capable of grouping data with irregular shapes [19]. DBSCAN has two main parameters, namely MinPts and Epsilon (ε). The epsilon parameter specifies the maximum distance to find the closest points, while minPts determines the least amount of points inside that distance to be considered one cluster [20].

Clustering steps using the DBSCAN algorithm [21] consist of the following:

1. Determine the MinPts and epsilon parameter values according to needs.
2. Choose a starting point p randomly as the reference point.
3. Compute the reachable epsilon distance based on the density of p by applying the Euclidean distance formula in Eq. (1).

$$d_{\{ij\}} = \sqrt{\sum_{a=1}^{p}(x_{\{ia\}} - x_{\{ja\}})^2} \quad (1)$$

The detailed explanation of the aforementioned formula is as follows:

d_{ij} = euclidean distance value
p = core point
$\sum$ = total sum result
a = feature index
x_{ia} = the value of variable a of object i
x_{ja} = the value of variable a of object j

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 6, Nomor 4, Juni 2025
e-ISSN 2685-998X
DOI 10.30865/json.v6i4

4. A cluster is formed when the number of points within the epsilon range exceeds the minPts value, with point p acting as the core point.
5. Recurrence steps 3 and 4 until all points have been processed. If p is a border point and no other points can be reached based on the density of p, then continue the process to the next point.

2.4 Evaluation

Evaluation is completed to examine the model's performance using measurements such as the Silhouette Score and Davies-Bouldin index to establish the ideal number of clusters. This process ensures the model's suitability to the expected results, significantly improving business understanding [22]. This study employs two evaluation metrics, specifically:

2.4.1 Silhouette Score

The first evaluation implemented to appraise the effectiveness of the clustering algorithm is the Silhouette Score. The Silhouette Score reflects how closely data points are grouped into the appropriate clusters. It falls within the ranges of -1 to 1, greater values approaching 1 suggest more distinct separation among clusters [23]. The Silhouette Score for the entire data set with a certain number of clusters is represented as sil(c), which is determined by computing the average silhouette accross all clusters [24] through the formula Eq. (2).

$$sil(c) = sil(k)\frac{1}{|k|}\sum_{i=1}^{k} sil(c_i) \quad (2)$$

The explanation of the formula as mentioned earlier is as follows:
$sil(c)$ = silhouette values.
$sil(k)$ = silhouette values of all clusters.
$|k|$ = the number of clusters k.
i = index of elements in the cluster.
$sil(c_i)$ = average silhouette value.

2.4.2 Davies Bouldin Index

The second metric used to assess the capability of the clustering model is the Davies-Bouldin Index. The Davies-Bouldin Index measures the separation quality of the cluster structure, where values closer to 0 indicate clusters that are more similar and well separated [25]. The Davies-Bouldin Index begins by calculating the SSW (Sum of Square Within Cluster) value [26], which is to find the cohesive matrix in the cluster, which is formulated in Eq. (3).

$$SSW_i = \frac{1}{m_i}\sum_{j=1}^{m_i} d(X_j, C_i) \quad (3)$$

The explanation of the aforementioned formula is as follows:
SSW = sum of square within cluster
m_i = amount of data in cluster i
j = index of data points in cluster i
$d(X_j, C_i)$ = distance from data i to cluster point i

Subsequently, Eq. (4) is applied to calculate the separation between the identified clusters.

$$SSB_{ij} = d(i,j) \quad (4)$$

The description of the formula mentioned above is presented as follows:
SSB = sum of square between cluster
$d(i,j)$ = euclidean distance of data i and data j

The succeeding step involves determining the proportion of the SSW and SSB calculations, as outlined in Eq. (5).

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (5)$$

The elucidation of the formula as mentioned above is detailed as follows:
R_{ij} = SSW and SSB calculation ratio

The final stage calculation to obtain the DBI value is as in equation 6.

$$DBI = \frac{1}{k}\sum_{i=1}^{k} i \neq j\ (R_{ij}) \quad (6)$$

The interpretation of the formula mentioned above is as follows:
DBI = Davies-Bouldin Index

3. RESULT AND DISCUSSION

The presentation of research findings follows a systematic and sequential approach, meticulously adhering to the methodological framework outlined in the study. Each stage is elaborated upon in detail to ensure comprehensive clarity regarding the analytical procedures undertaken, thereby enhancing transparency and reliability in interpreting results. Beginning with data collection, processing through data preprocessing, and culminating in the clustering analysis using the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm,

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 6, Nomor 4, Juni 2025
e-ISSN 2685-998X
DOI 10.30865/json.v6i4

every phase is highlighted for its substantive contribution to the outcomes. The evaluation of clustering results constitutes a pivotal element of the presentation, with metrics such as the Silhouette Score and Davies-Bouldin Index serving as critical tools for validating the effectiveness and robustness of the clustering methodology. This systematic exposition aims to furnish novel insights that resonate with the overarching research objectives and to align closely with the complexities of the problem under investigation. Further elaboration on these findings will be addressed comprehensively in the following section.

3.1 Data Collection

Data collection is a crucial preliminary phase in this study, laying the foundation for subsequent analytical stages. The dataset employed was sourced from credible and authoritative references to ensure its validity, reliability, and relevance. Specifically, it was obtained from the Central Statistics Agency (BPS) official platform of Central Java Province, which provides comprehensive information on poverty levels across the region for the year 2023. The dataset comprises 35 entries, each corresponding to a district or city, and includes three essential attributes: UMK (District Minimum Wage), unemployment rate, and district/city identification. The data was manually recorded in a spreadsheet format to facilitate effective preprocessing and clustering analysis, ensure accuracy, and streamline the initial preparation phase. The spreadsheet was then converted into a CSV format to enhance compatibility with analytical tools and optimize data processing efficiency. The meticulously curated dataset is presented in Table 2, forming the basis for a robust and transparent analysis framework that aligns with the research objectives. Further elaboration on its utilization will be discussed in the subsequent sections.

Table 2. Poverty Dataset

District/City	District Minimum Wage	Unemployment Rate
Banjarnegara	2038005	6.26
Banyumas	2118125	6.35
Batang	2284627	6.06
Blora	2040080	3.10
Boyolali	2155712	4.05
Brebes	2018000	8.98
Cilacap	2383090	8.74
Demak	2680421	5.38
Grobogan	2029569	4.02
Jepara	2272627	3.35
Karanganyar	2207484	4.35
Kebumen	2043902	5.11
Kendal	2508300	5.76
Klaten	2152323	4.20
Kota Magelang	2066007	5.25
Kota Pekalongan	2305283	5.02
Kota Salatiga	2284179	4.57
Kota Semarang	3060349	5.99
Kota Surakarta	2197169	4.58
Kota Tegal	2145012	6.05
Kudus	2439813	3.25
Magelang	2236777	4.42
Pati	2107697	4.29
Pekalongan	2334886	3.25
Pemalang	2091783	6.55
Purbalingga	2130981	5.61
Purworejo	2043902	4.02
Rembang	2015927	2.60
Kab. Semarang	2480988	5.99
Sragen	1969569	3.87
Sukoharjo	2138248	3.40
Tegal	2106238	8.60
Temanggung	2027569	2.32
Wonogiri	1968448	1.92
Wonosobo	2076209	4.95

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 6, Nomor 4, Juni 2025
e-ISSN 2685-998X
DOI 10.30865/json.v6i4

3.2 Data Preprocessing

3.2.1 Data Cleaning and Outlier Handling

The data cleaning stage is essential to ensure the dataset is well-prepared and optimized for subsequent analytical steps. This study identified the UMK (District Minimum Wage) attribute as containing outliers, as evidenced by two data points positioned outside the whisker boundaries in the attribute's boxplot. These anomalies highlight the presence of extreme values within the dataset that could skew the outcomes of the clustering analysis. To address this, the study focused on managing outliers applying the Interquartile Range (IQR) method, which measures data dispersion between the first quartile (Q1) and third quartile (Q3). This method has proven effective for detecting and mitigating the impact of extreme values unrepresentative of the broader dataset.

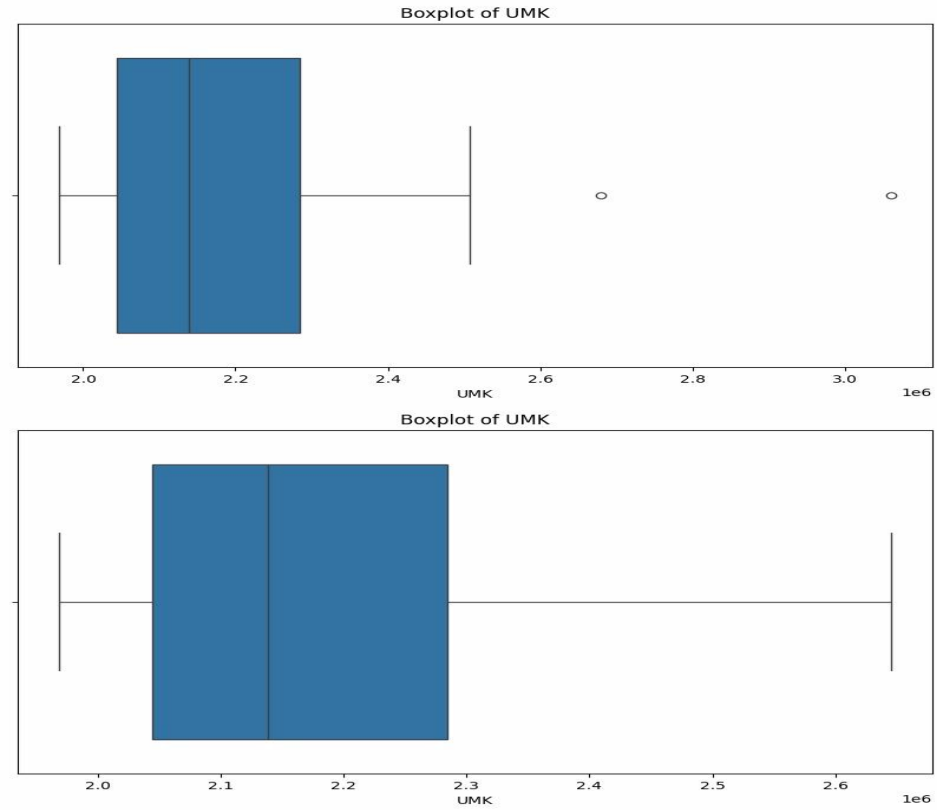


Figure 2. Boxplot Handling Outliers

Outlier detection and handling were further supported by boxplot visualization, offering a precise comparative analysis of the data's distribution before and after the cleaning process. This visualization not only facilitates the identification of significant shifts in data structure but also ensures transparency in demonstrating the effectiveness of the outlier management approach. The dataset becomes more coherent and reliable by eliminating or adjusting outliers, thereby strengthening the foundation for robust clustering analysis. Figure 2 provides detailed insights into the outlier identification process and the subsequent data adjustments performed in this study. This rigorous methodology underscores the study's commitment to data integrity and analytical precision, paving the way for accurate and meaningful results.

3.2.2 Data Transformation

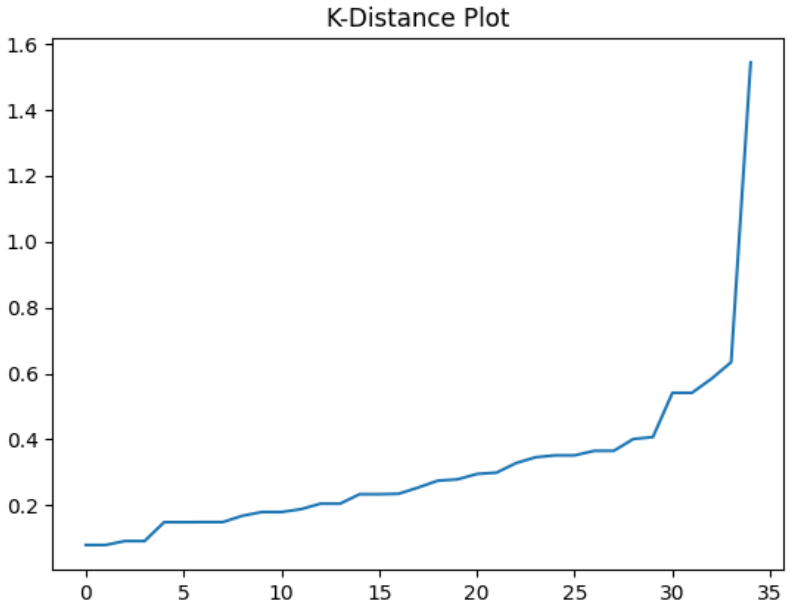
Once the data cleaning process is completed, the next phase is transformation using the standardization method of Python programming through the scikit-learn library and the StandardScaler module. This transformation changes the data to follow a uniform distribution with a mean of zero and a standard deviation of one to overcome the scale differences between variables. This step ensures that the contribution of each variable is equal in the analysis without the dominance of large-scale features. The data transformation results, which aim to improve the balance and accuracy of the analysis, are presented in detail in Table 3 to provide an overview of the changes after this process.

Table 3. Poverty Data Transformation

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 6, Nomor 4, Juni 2025
e-ISSN 2685-998X
DOI 10.30865/json.v6i4

No	District Minimum Wage	Unemployment Rate
0	-0.854040	0.802490
1	-0.407998	0.856400
2	0.518955	0.682690
3	-0.842488	-1.090352
4	-0.198742	-0.521302

3.3 DBSCAN Clustering



The clustering process in this study was carried out using Python programming with Google Collaboratory tools. The library used in this process is scikit-learn with the DBSCAN module. The attributes used are UMK (district minimum wage) and unemployment rate, where the clustering results will then be labeled based on the district/city. Assessment of the epsilon (ε) and minPts values is done by utilizing the k-distance plot to establish the optimal cluster count. Based on research [27], the k-distance plot is a valuable tool to help determine the epsilon (ε) parameter value in the DBSCAN algorithm by visualizing the range of each data point to its nearest neighbor (K-Nearest Neighbor).

Figure 3. K-Distance Plot

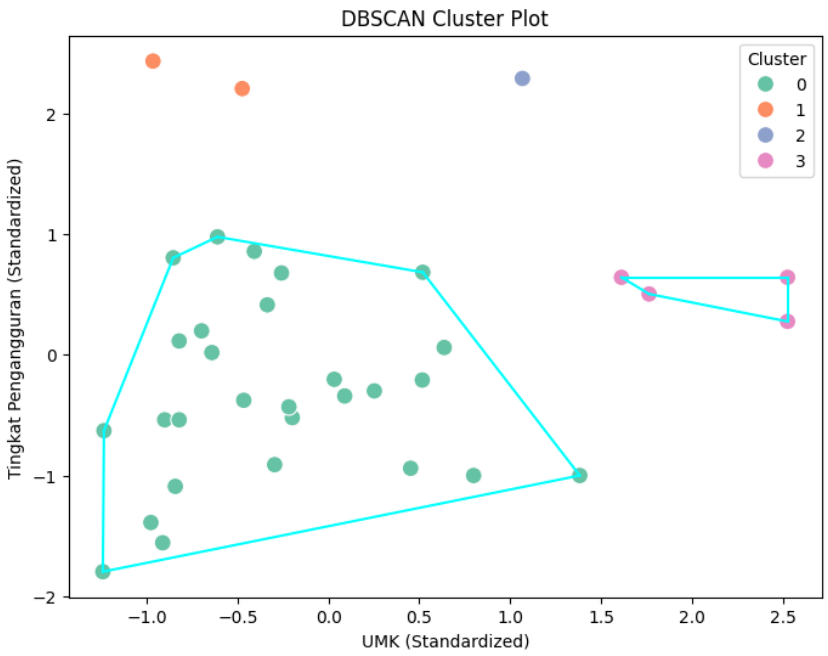


Figure 4. DBSCAN Clustering

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 6, Nomor 4, Juni 2025
e-ISSN 2685-998X
DOI 10.30865/json.v6i4

Table 4. Members of the Regency/City Cluster

Cluster	Member
Cluster 0	Banjarnegara, Banyumas, Batang, Blora, Boyolali, Grobogan, Jepara, Karanganyar, Kebumen, Klaten, Kota Magelang, Kota Pekalongan, Kota Salatiga, Kota Surakarta, Kota Tegal, Kudus, Magelang, Pati, Pekalongan, Pemalang, Purbalingga, Purworejo, Rembang, Sragen, Sukoharjo, Temanggung, Wonogiri, and Wonosobo
Cluster 1	Brebes and Tegal
Cluster 2	Cilacap
Cluster 3	Demak, Kendal, Kab. Semarang, and Kota Semarang

In Fig. 3, the epsilon and minPts values chosen for DBSCAN clustering are 0.8 and 1, resulting in four clusters. The minPts value is set at 1 to ensure no clusters interfere with the analysis results. The resulting cluster visualization is shown in Figure 4. Cluster membership details are provided in Table 4.

The clustering results indicate that Cluster 0 comprises 28 districts/cities in Central Java Province, Cluster 1 contains 2 districts/cities, Cluster 2 includes 1 district/city, and Cluster 3 contains 4 districts/cities. A summary of the characteristics of each cluster is presented in Table 5.

The interpretation of clustering results is as follows :

a. Cluster 0 describes areas with moderate poverty categories. Despite lower unemployment rates, low district minimum wages (UMK) still reflect significant economic disparities.
b. Cluster 1 is an area with a high poverty category, where the unemployment rate and the UMK are low, thus causing quite significant economic and social pressures.
c. Cluster 2 shows areas with very high poverty categories. Despite having the highest UMK, this area also has a high unemployment rate.
d. Cluster 3 describes areas with a low poverty category. This area has a fairly high UMK and a moderate unemployment rate, reflecting relatively better socio-economic conditions than other clusters' regions.

Table 5. Characteristic of Each Cluster

Cluster	Count	District Minimum Wage (mean)	District Minimum Wage (std)	Unemployment Rate (mean)	Unemployment Rate (std)
0	28	2139730.491786	119314.918089	4.384643	1.242705
1	2	2062119.00000	62393.688158	8.790000	0.268701
2	1	2383090.0000	NaN	8.740000	NaN
3	4	2569899.388750	2569899.388750	5.780000	0.287866

3.4 Evaluation

3.4.1 Silhouette Score

The evaluation of clustering performance in this study was carried out using the Silhouette Score metric. Considering the evaluation outcome, the Silhouette Score value obtained was 0.443 from the four clusters formed. This value indicates that the clustering process has succeeded in forming well-defined clusters. The higher the Silhouette Score value, the better the clustering quality; this reflects that the data in the cluster has high similarity while the distance from the data in other clusters is quite far. Although the value is not too high, this result shows that the cluster separation can be accepted relatively well.

3.4.2 Davies Bouldin Index

This study's clustering performance was also assessed with the Davies-Bouldin Index. Based on the outcome of the clustering process conducted, the Davies-Bouldin Index value obtained was 0.441. This value is used to evaluate the level of separation between clusters and how close the distance between clusters is. This low Davies-Bouldin Index value indicates that the separation between clusters is quite clear, and the clusters formed have good separation quality.

4. CONCLUSION

Based on the research that has been done, the clustering technique for grouping poverty areas in Central Java Province using the DBSCAN algorithm is quite good and provides clear insights. The DBSCAN algorithm

produces 4 clusters with details of cluster 0 containing 28 districts/cities, cluster 1 containing 2 districts/cities, cluster 2 containing 1 district/city, and cluster 3 containing 4 districts/cities. The areas included in cluster 0 are areas with moderate poverty levels. Although the unemployment rate is low, the UMK (District Minimum Wage) is lower, causing significant economic disparities. The regions in cluster 1 have high poverty levels, low UMK (district minimum wage), and high unemployment rates, reflecting significant socio-economic pressures. The areas included in cluster 2 are areas with very high poverty rates because the district minimum wage (UMK) is high, but the unemployment rate is also very high. The areas included in Cluster 3 are prosperous areas with low poverty rates, higher UMK (district minimum wage), and lower unemployment rates. The clustering process uses 2 metric evaluations, namely the Silhouette Score and the Davies-Bouldin Index, which produce good values, 0.443 and 0.441, so the clustering results are promising. Thus, the outcomes of this research are capable to cluster poverty levels based on districts/cities in Central Java province. Further analysis can be conducted using different algorithms and variables to compare the results.

REFERENCES

[1] A. Bahauddin, A. Fatmawati, and F. Permata Sari, "Analisis Clustering Provinsi Di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means," *J. Manaj. Inform. dan Sist. Inf.*, vol. 4, no. 1, pp. 1–8, 2021, doi: 10.36595/misi.v4i1.216.

[2] L. S. Hasibuan, "Analisis Pengaruh Ipm, Inflasi, Pertumbuhan Ekonomi Terhadap Pengangguran Dan Kemiskinan Di Indonesia," *J. Penelit. Pendidik. Sos. Hum.*, vol. 8, no. 1, pp. 53–62, 2023, [Online]. Available: https://jurnal-lp2m.umnaw.ac.id/index.php/JP2SH/article/view/2075/1261

[3] U. Qulsum, J. Anggraini, and E. Putri, "ANALYSIS OF POVERTY AS A MAIN PROBLEM OF THE INDONESIAN," 2024.

[4] A. A. M. Davidescu, T. M. Nae, and M. S. Florescu, "From Policy to Impact: Advancing Economic Development and Tackling Social Inequities in Central and Eastern Europe," *Economies*, vol. 12, no. 2, 2024, doi: 10.3390/economies12020028.

[5] N. N. F. R, D. S. Anggraeni, and U. Enri, "Pengelompokkan Data Kemiskinan Provinsi Jawa Barat Menggunakan Algoritma K-Means dengan Silhouette Coefficient," *Tematik*, vol. 9, no. 1, pp. 29–35, 2022, doi: 10.38204/tematik.v9i1.901.

[6] N. K. Surbakti, "Data Mining Pengelompokan Pasien Rawat Inap Peserta BPJS Menggunakan Metode Clustering (Studi Kasus : RSU.Bangkatan)," *J. Inf. Technol.*, vol. 1, no. 2, pp. 47–53, 2021, doi: 10.32938/jitu.v1i2.1470.

[7] M. Richia Putri, Gibran Satya Nugraha, and Ramaditia Dwiyansaputra, "Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan Menggunakan Metode K-Means Clustering," *J. Comput. Sci. Informatics Eng.*, vol. 7, no. 1, pp. 1–8, 2023, doi: 10.29303/jcosine.v7i1.509.

[8] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024, [Online]. Available: https://doi.org/10.58602/jima-ilkom.v3i1.26

[9] R. R. A. Rahman and A. W. Wijayanto, "Pengelompokan Data Gempa Bumi Menggunakan Algoritma Dbscan," *J. Meteorol. dan Geofis.*, vol. 22, no. 1, p. 31, 2021, doi: 10.31172/jmg.v22i1.738.

[10] N. Sukarno Wijaya, M. Jajuli, and B. Arif Dermawan, "Penerapan Algoritma K-Means Clustering Dalam Menentukan Daerah Prioritas Penanganan Kemiskinan Di Wilayah Jawa Timur," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 4, pp. 7579–7584, 2024, doi: 10.36040/jati.v8i4.10248.

[11] S. N. Mayasari and J. Nugraha, "Implementasi K-Means Cluster Analysis untuk Mengelompokkan Kabupaten/Kota Berdasarkan Data Kemiskinan di Provinsi Jawa Tengah Tahun 2022," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 317–329, 2023, doi: 10.24002/konstelasi.v3i2.7200.

[12] F. Zahra, A. Khalif, and B. N. Sari, "Pengelompokan Tingkat Kemiskinan di Setiap Provinsi di Indonesia Menggunakan Algoritma K-Medoids," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 2830–7062, 2024, [Online]. Available: http://dx.doi.org/10.23960/jitet.v12i2.4199

[13] D. J. Manalu, R. Rahmawati, and T. Widiharih, "PENGELOMPOKAN TITIK GEMPA DI PULAU SULAWESI MENGGUNAKAN ALGORITMA ST-DBSCAN (Spatio Temporal-Density Based Spatial Clustering Application with Noise)," *J. Gaussian*, vol. 10, no. 4, pp. 554–561, 2021, doi: 10.14710/j.gauss.v10i4.29499.

[14] A. F. Boer, M. Al Haris, and R. Wasono, "Pengelompokan Daerah Rawan Bencana di Pulau Sumatera dengan Metode Density Based Spatial Clustering of Applications with Noise (DBSCAN)," *Semin. Nas. Unimus*, pp. 389–400, 2023, [Online]. Available: https://prosiding.unimus.ac.id/index.php/semnas/article/download/1481/1485

[15] Purwanty, S. Martha, and S. Aprizkiyandari, "Pengelompokan kabupaten/kota di kalimantan barat berdasarkan faktor-faktor kemiskinan menggunakan metode dbscan," *Bul. Ilm. Mat. Stat. dan Ter.*, vol. 13, no. 1, pp. 17–26, 2024.

[16] A. A. Bushra and G. Yi, "Comparative Analysis Review of Pioneering DBSCAN and Successive Density-Based Clustering Algorithms," *IEEE Access*, vol. 9, pp. 87918–87935, 2021, doi: 10.1109/ACCESS.2021.3089036.

[17] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.

[18] F. Dwi Handoko, A. Fauzi, D. Ryan, F. Kurniasih, P. Mutiara, and S. Taqwaning Afifi, "Transformasi Data Menjadi Informasi Pada Bisnis Intelijen," *J. Ilmu Hukum, Hum. dan Polit.*, vol. 2, no. 3, pp. 313–19, 2022, doi: 10.38035/jihhp.v2i3.1043.

[19] M. S. Hasibuan, A. H. Lubis, and M. N. Sari, "Perbandingan algoritma clustering dbscan dan k-means dalam pengelompokan siswa terbaik Comparison of dbscan and k-means clustering algorithms in grouping the best students," vol. 5, pp. 301–309, 2024.

[20] R. A. Satria, I. Indriati, and S. Sutrisno, "Pengelompokan Hasil Pencarian Skripsi Berbahasa Indonesia Menggunakan Metode DBSCAN dengan Pembobotan BM25," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 4, pp. 781–790, 2023, doi: 10.25126/jtiik.20241046899.

[21] 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

[22] M. Sholeh, S. Suraya, and D. Andayati, "Penerapan Data Mining pada Model Clustering Data Kuesioner Mahasiswa terhadap Kinerja Dosen," *J. Eksplora Inform.*, vol. 13, no. 2, pp. 208–217, 2024, doi: 10.30864/eksplora.v13i2.751.

[23] M. Daffa Rachman and A. Voutama, "Implementasi Algoritma K-Means Dalam Sistem Rekomendasi Musik Menggunakan Python," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 3857–3862, 2024, doi: 10.36040/jati.v8i3.9635.

[24] Y. I. Kurniawan, "Pengelompokan Prioritas Negara Yang Membutuhkan Bantuan Menggunakan Clustering K-Means dengan Elbow dan Silhouette," *J. Pendidik. dan Teknol. Indones.*, vol. 3, no. 10, pp. 455–463, 2023, [Online]. Available: https://doi.org/10.52436/1.jpti.343

[25] S. F. Djun, I. G. A. Gunadi, and S. Sariyasa, "Analisis Segmentasi Pelanggan pada Bisnis dengan Menggunakan Metode K-Means Clustering pada Model Data RFM," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 5, no. 4, pp. 354–364, 2024, doi: 10.35746/jtim.v5i4.434.

[26] F. Fathurrahman, S. Harini, and R. Kusumawati, "Evaluasi Clustering K-Means Dan K-Medoid Pada Persebaran Covid-19 Di Indonesia Dengan Metode Davies-Bouldin Index (Dbi)," *J. Mnemon.*, vol. 6, no. 2, pp. 117–128, 2023, doi: 10.36040/mnemonic.v6i2.6642.

[27] Nurmadani and F. Rakhmawati, "Clustering flood prone areas in deli serdang regency using density-based spatial clustering of applications with noise (dbscan) method," vol. 6, no. 2, pp. 261–272, 2023, doi: 10.24042/djm.