Digital Zone
Jurnal Teknologi Informasi dan Komunikasi

Volume 17 Issue 1 Year 2026 | Page 68-79 | e-ISSN: 2477-3255 | ISSN: 2086-4884
Received: 04-04-2026 | Revised: 09-04-2026 | Accepted: 02-05-2026

Modeling Multidimensional Social Welfare Using Machine Learning Analysis

Rima Liana Gema[1]*, Mutiana Pratiwi[2], Silky Safira[3]
[1,2,3] Universitas Putra Indonesia YPTK Padang, Indonesia
Email: rimalianagema@gmail.com, mutianapratiwi@upiyptk.ac.id, silkysafira@upiyptk.ac.id
*Corespondence E-mail: rimalianagema@gmail.com

Abstract: *The measurement of regional welfare in Indonesia is still dominated by economic indicators and therefore does not fully capture broader social conditions. This study aims to model and analyze multidimensional social welfare across provinces in Indonesia using a machine learning approach. The analysis is based on official data from Statistics Indonesia (BPS), covering five key dimensions: health, economy, education, infrastructure, and employment. The data were transformed using Benefit–Cost normalization and analyzed using the K-Means clustering algorithm. The results reveal four distinct welfare clusters with varying characteristics. The highest welfare cluster achieves an average index value of 0.6143, while the medium and lower-middle clusters record values of 0.4673 and 0.4210, respectively. The lowest welfare cluster has an index value of 0.3468, indicating limitations in health, economic, and employment dimensions. These findings highlight significant disparities in regional welfare conditions and demonstrate the importance of a multidimensional analytical approach. This study contributes by integrating multidimensional welfare indicators with a machine learning clustering method to provide a more objective and data-driven classification of regional welfare.*

Keywords: *social welfare, multidimensional analysis, machine learning, benefit–cost, K-Means*

1. Introduction

Social welfare holds a strategic position as a national development goal, as it serves as a measure of the government's success in improving the quality of life for the community [1]. Welfare is not simply understood as increasing regional income, but also as a development process that encourages the achievement of a better quality of life. This condition indicates that community welfare is not solely shaped by economic factors but is also influenced by the fulfillment of other social needs, both material and non-material. Therefore, the concept of welfare must be examined holistically as a multidimensional phenomenon [2].

Measuring welfare requires indicators that comprehensively describe the social conditions of the community. Using economic indicators alone is not sufficient, as welfare is also reflected in the conditions of education, health, employment, demographics, and other social aspects related to quality of life [3]. Based on this perspective, this study assesses regional welfare through five main dimensions that represent the social conditions of the community: economics, education, health, basic infrastructure, and workforce involvement.

Despite ongoing development, the level of welfare in Indonesia is not evenly distributed across regions [4]. Disparities in access to education, health services, employment, and infrastructure have left some regions lagging behind in achieving welfare. Realities such as the persistently high prevalence of stunting in certain regions demonstrate that economic growth is not always aligned with improvements in quality of life [5]. The main problem in measuring

https://doi.org/10.31849/digitalzone.v17i1. 32730

regional welfare is the dominance of assessments based on single economic indicators, which fail to reflect the inequalities in quality of life determined by the dimensions of education, health, infrastructure, and employment opportunities. Therefore, a welfare evaluation model capable of mapping social characteristics multidimensionally is needed.

To produce a more objective evaluation, welfare analysis can be conducted using a machine learning approach that analyzes data patterns without being influenced by subjective interpretations of researchers. One form of machine learning method is unsupervised learning, which works by identifying data structures without requiring category labels [6], [7]. Through this approach, regions can be grouped based on shared characteristics of their welfare indicators. Of the various unsupervised learning methods, the K-Means algorithm is one of the most widely used clustering techniques in social analysis [8]. This algorithm divides data into specific clusters by establishing cluster centers (centroids) that are gradually updated until optimal clustering is achieved [9]. This procedure aligns with the basic principles of clustering, namely placing similar objects in one group [10] and separating them from objects with different characteristics [11]. In the context of regional welfare mapping, the K-Means algorithm can automatically identify regions with similar socioeconomic profiles, ensuring that clustering results are not influenced by researcher bias. To ensure that all dimensions of welfare are given equal weight in the cluster formation process, the data is first normalized so that each variable is on a proportional comparison scale. This process makes regional classification more objective, accurate, and able to reflect multidimensional welfare conditions [12]. Several previous studies have applied clustering and machine learning approaches to analyze social welfare. For instance, [13] and [14] utilized the K-Means algorithm to classify regional welfare; however, their studies were limited to specific regions and did not fully incorporate balanced multidimensional indicators. Similarly, [15] compared clustering algorithms but focused primarily on economic indicators without addressing broader welfare dimensions.

Other studies, such as [16] and [17], focused mainly on poverty-related indicators, while [12] conducted analysis at a micro-level, limiting its representation of regional welfare conditions. In addition, [18] applied a machine learning approach to analyze multidimensional poverty; however, the study focused on classification rather than clustering, which limits its ability to identify natural groupings of regions. Previous studies have also highlighted education as an important determinant of welfare; however, these studies often analyze it as a single variable rather than integrating it into a multidimensional framework

These limitations indicate that previous studies have not fully addressed the need for a comprehensive, multidimensional, and balanced approach to welfare classification. Therefore, this study proposes a multidimensional welfare analysis by integrating five key dimensions health, economy, education, infrastructure, and employment combined with Benefit Cost normalization to ensure proportional contribution of each indicator. This study utilizes official data from the Central Statistics Agency (BPS, 2023) and stunting information from the National Data Portal to model regional welfare levels using the K-Means algorithm. The results are expected to provide a more objective and data-driven classification of welfare and support more targeted regional development policies.

In addition, this study differs from previous research by integrating a Benefit–Cost normalization approach within a multidimensional clustering framework to ensure that each welfare indicator contributes proportionally to the analysis. Unlike earlier studies that focus on limited indicators or localized regions, this research utilizes nationwide data covering all provinces in Indonesia, providing a more comprehensive and scalable analysis.

Furthermore, this study not only applies K-Means clustering but also evaluates clustering quality using both the Elbow Method and Silhouette Score to ensure robustness and validity of the results. Therefore, the main contribution of this study lies in providing a more objective, data-driven, and multidimensional framework for regional welfare classification that can better support evidence-based policymaking.

2. Research Method

This study employs a quantitative approach using an unsupervised machine learning technique, specifically the K-Means algorithm, to map variations in social welfare levels across provinces in Indonesia based on multiple dimensions. This method is selected due to its ability to cluster data with similar characteristics without requiring predefined labels. The research framework begins with the data collection stage, where secondary data are obtained from official government sources. This is followed by data preprocessing and transformation, which includes scaling the data into a 0–1 range, determining indicator directions (benefit or cost), and applying benefit–cost normalization to ensure comparability across variables. Subsequently, the clustering process is performed using the K-Means algorithm. The optimal number of clusters is determined using the Elbow method and further validated using the Silhouette Score to ensure clustering quality.

The process is conducted iteratively until a stable cluster structure is achieved. Finally, the analysis and interpretation stage is carried out to identify the characteristics of each cluster based on the selected social welfare indicators. Overall, the research workflow consists of the following stages: data collection, data preprocessing and transformation, K-Means clustering, evaluation (Elbow and Silhouette methods), and result interpretation, as illustrated in Figure 1.

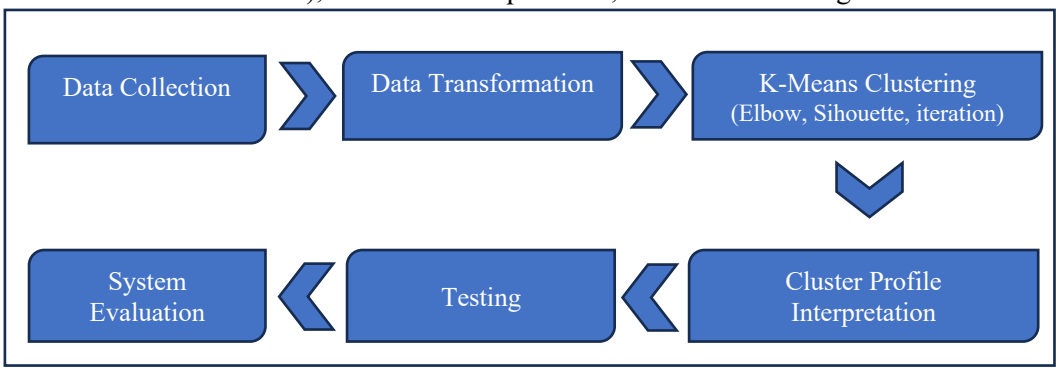


Figure 1. Flow of Research Stages

2.1 Data Collection

The initial dataset of this study consists of six welfare indicators across 34 provinces in Indonesia, representing aspects of health, economy, education, infrastructure, and employment. The distribution of values for each indicator shows considerable variation across regions. For instance, the prevalence of stunting ranges from 8.00 to 35.29 percent, the poverty rate varies between 3.80 and 21.09 percent, and per capita expenditure ranges from IDR 975,854 to IDR 2,794,485. Similar disparities are observed in the average years of schooling, access to safe drinking water, and unemployment rates. These variations highlight significant differences in welfare conditions among provinces, indicating that a multidimensional modeling approach is necessary to produce a more comprehensive and objective mapping. The complete dataset used in this study is presented in Table 2.

Table 1. Dataset of Welfare Indicators Before Transformation

No	Dimension	Health	Economiy		Education	Infrastru cture	Work Force
	Province	Stunting (%)	Poverty Rate (%)	Per Capita Expenditure (IDR/month)	Average Years of Schooling (years)	Access to Safe Drinkin g Water (%)	Unemploy ment Rate(%)
		Cost	Cost	Benefit	Benefit	Benefit	Cost

1	Aceh	31.20	12.64	1264733	9.95	90.08	5.75
2	Sumatera Utara	21.20	7.19	1340957	10.18	92.94	5.60
3	Sumatera Barat	25.20	5.42	1480853	9.72	86.85	5.75
4	Riau	17.00	6.36	1563165	9.69	92.24	3.70
5	Kep. Riau	15.39	4.78	2109071	10.65	94.71	6.39
6	Jambi	18.00	7.26	1472501	9.26	82.16	4.48
7	Sumatera Selatan	18.60	10.51	1289001	8.98	87.23	3.86
8	Kep. Bangka Belitung	18.50	5.08	1793074	8.78	82.97	4.63
9	Bengkulu	19.79	12.52	1415976	9.40	72.10	3.11
10	Lampung	15.19	10.62	1201473	8.80	84.51	4.19
11	DKI Jakarta	14.80	4.14	2794485	11.49	99.96	6.21
12	Jawa Barat	20.20	7.08	1633032	9.24	94.90	6.75
13	Jawa Tengah	20.79	9.58	1271678	8.47	95.43	4.78
14	DIY	16.39	10.40	1758865	10.23	96.91	3.48
15	Jawa Timur	19.20	9.56	1356943	8.69	96.93	4.19
16	Banten	20.00	5.70	1772162	9.55	94.01	6.68
17	Kalimantan Barat	27.79	6.25	1388077	8.28	82.13	4.86
18	Kalimantan Tengah	26.89	5.26	1590691	9.17	78.71	4.01
19	Kalimantan Selatan	24.60	4.02	1534039	9.02	77.34	4.20
20	Kalimantan Timur	23.89	5.51	2042791	10.22	89.11	5.14
21	Kalimantan Utara	22.10	5.38	1658042	9.58	89.42	3.90
22	Sulawesi Utara	20.50	6.70	1382429	10.02	94.46	5.85
23	Gorontalo	23.79	13.87	1262268	8.64	96.13	3.13
24	Sulawesi Tengah	28.20	11.04	1231453	9.28	87.86	2.94
25	Sulawesi Barat	35.00	10.71	1109589	8.56	80.14	2.68
26	Sulawesi Selatan	27.20	7.77	1283081	9.22	92.23	4.19
27	Sulawesi Tenggara	27.70	10.63	1227163	9.74	95.31	3.09
28	Bali	8.00	3.80	1872760	9.87	98.32	1.79
29	NTB	32.70	11.91	1277139	8.50	96.13	2.73
30	NTT	35.29	19.02	975854	8.48	88.55	3.02
31	Maluku	26.10	15.78	1277569	10.44	93.32	6.11
32	Maluku Utara	26.10	6.03	1447310	9.70	88.92	4.03
33	Papua	34.59	18.09	1578894	10.70	85.84	6.48
34	Papua Barat	30.00	21.09	1643488	9.87	82.38	4.13

2.2 Data Transformation

At the data transformation stage, the dataset is processed to ensure that each variable contributes proportionally in the clustering process. This stage includes three main steps: scaling standardization, adjustment of indicator direction (benefit and cost), and normalization using the Benefit–Cost approach. These processes are applied sequentially to prepare the data for further analysis. The steps involved are as follows:

1. Scaling Standardization (0–1 Scaling)

Each indicator has different scales and units; therefore, Min–Max Scaling is applied to transform all variables into a uniform range between 0 and 1 [19].

The normalization is calculated as follows:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

2. Indicator Direction Adjustment (Benefit–Cost)

Some indicators represent cost attributes of welfare, such as poverty, stunting, and unemployment. To ensure consistency in interpretation, cost indicators are converted into benefit indicators using the following transformation:

$$X' = 1 - X \quad (2)$$

3. Benefit–Cost Normalization

This process ensures that all indicators have the same direction (higher values indicate better welfare conditions) and are measured on a comparable scale, allowing fair contribution in the clustering process.

2.3 K-Means Clustering

After the data transformation stage, clustering is performed using the K-Means algorithm. The optimal number of clusters is determined using the Elbow Method and Silhouette Score to ensure the stability and quality of clustering results [20]. The algorithm initializes centroids randomly, assigns each province to the nearest cluster based on Euclidean distance, and updates centroid positions iteratively until convergence is achieved. This iterative process results in four optimal clusters.

2.4 Testing

At this stage, the clustering results are tested to ensure consistency and stability. The testing process evaluates whether the formed clusters remain stable across iterations and whether the grouping reflects meaningful patterns in the data. This step ensures that the clustering output is reliable before further evaluation.

2.5 System Evaluation

System evaluation is conducted using quantitative metrics, including the Elbow Method and Silhouette Score, to assess clustering performance. The Elbow Method is used to identify the optimal number of clusters based on the Within-Cluster Sum of Squares (WCSS), while the Silhouette Score evaluates the separation and cohesion of clusters. These evaluation results confirm the validity of the clustering structure.

2.6 Cluster Profile Interpretation

The final stage involves interpreting the clustering results. Each cluster is analyzed based on six welfare indicators to identify the characteristics of provinces with similar profiles. The interpretation includes calculating the average value of indicators within each cluster, comparing inter-cluster differences, and constructing welfare profiles ranging from high to low categories. Additionally, each cluster is evaluated to identify regional strengths, challenges, and disparities.

This analysis provides a comprehensive understanding of multidimensional social welfare conditions across provinces in Indonesia.

2.7 Implementation Details

The implementation of this study was carried out using the Python programming language, supported by several libraries including NumPy and Pandas for data processing, Scikit-learn for machine learning modeling, and Matplotlib for data visualization. The K-Means clustering algorithm was applied using the KMeans function from the Scikit-learn library, with the number of clusters (k) set to 4 based on the results of the Elbow Method and Silhouette Score evaluation. In this study, the K-Means algorithm was configured using the k-means++ initialization method to improve the stability of centroid selection. The maximum number of iterations was set to 300, with 10 different initial centroid configurations (n_init = 10) to avoid local optima. In addition, a fixed random state value of 42 was applied to ensure that the clustering results are consistent and reproducible across multiple runs. Prior to clustering, all variables were preprocessed using Min–Max scaling and Benefit–Cost normalization to standardize the range and direction of each indicator. This preprocessing step ensures that all variables contribute proportionally to the clustering process and prevents bias caused by differences in scale. Furthermore, the dataset used in this study was obtained from publicly available sources, enabling transparency and allowing future researchers to replicate and validate the findings under the same computational conditions.

3. Results and Discussion

3.1 Results

This section presents the main findings of the multidimensional social welfare analysis conducted across 34 provinces in Indonesia. The results are obtained through data transformation followed by clustering using the K-Means algorithm. The identified clusters are described based on the patterns of welfare indicators to highlight similarities in provincial characteristics.

3.1.1 Overview of Research Data

The initial dataset consists of six welfare indicators across 34 provinces in Indonesia, representing the dimensions of health, economy, education, infrastructure, and employment. The distribution of values across these indicators shows significant variation among regions. For instance, the prevalence of stunting ranges from 8.00 to 35.29 percent, while the poverty rate varies between 3.80 and 21.09 percent. In addition, per capita expenditure ranges from IDR 975,854 to IDR 2,794,485. Similar disparities are also observed in the average years of schooling, access to safe drinking water, and unemployment rate. These variations indicate substantial differences in welfare conditions across provinces, emphasizing the need for a multidimensional modeling approach to provide a more comprehensive and objective mapping of social welfare.

3.1.2 Transformation

Data transformation was performed to standardize the scale and direction of all welfare indicators so that they could be analyzed proportionally using the K-Means algorithm. This process included Min–Max normalization to equalize indicator values into the 0–1 range as shown in Formula (1), as well as converting the direction of cost indicators to benefits using the inverse transformation according to Formula (2). Through this stage, all indicators have a consistent interpretation, namely that higher values indicate better welfare conditions. In line with this process, several cost indicators also underwent naming adjustments, namely Stunting (%) was changed to Stunting Prevention, Poor Population (%) to Poverty Rate (Inverse), and Unemployment Rate (%) to Employment Engagement Index, while other indicators retained their original names. The final data transformation results are presented in Table 3 and used as the basis for the clustering process.

Table 2. Result Transformation of Welfare Indicators (Scale 0–1, Benefit Direction)

No	Dimension	Health	Economy		Education	Infrastructure	Work Force
	Province	Stunting Prevention	Poverty Rate (Inverse)	Per Capita Expenditure	Average Years of Schooling	Access to Safe Drinking Water	Employment Engagement Index
1	Aceh	0.1499	0.4887	0.1588	0.5202	0.6454	0.2016
2	Sumatera Utara	0.5163	0.8039	0.2008	0.5919	0.7480	0.2319
3	Sumatera Barat	0.3697	0.9063	0.2777	0.4486	0.5294	0.2016
4	Riau	0.6702	0.8519	0.3229	0.4393	0.7229	0.6149
5	Kep. Riau	0.7292	0.9433	0.6231	0.7383	0.8116	0.0726
6	Jambi	0.6336	0.7999	0.2731	0.3053	0.3611	0.4577
7	Sumatera Selatan	0.6116	0.6119	0.1722	0.2181	0.5431	0.5827
8	Kep. Bangka Belitung	0.6152	0.9260	0.4494	0.1558	0.3902	0.4274
9	Bengkulu	0.5680	0.4957	0.2420	0.3489	0.0000	0.7339
10	Lampung	0.7365	0.6056	0.1241	0.1620	0.4454	0.5161
11	DKI Jakarta	0.7508	0.9803	1.0000	1.0000	1.0000	0.1089
12	Jawa Barat	0.5529	0.8103	0.3614	0.2991	0.8184	0.0000
13	Jawa Tengah	0.5313	0.6657	0.1627	0.0592	0.8374	0.3972
14	DIY	0.6926	0.6183	0.4305	0.6075	0.8905	0.6593
15	Jawa Timur	0.5896	0.6669	0.2095	0.1277	0.8912	0.5161
16	Banten	0.5603	0.8901	0.4379	0.3956	0.7864	0.0141
17	Kalimantan Barat	0.2748	0.8583	0.2267	0.0000	0.3600	0.3810
18	Kalimantan Tengah	0.3078	0.9156	0.3381	0.2773	0.2373	0.5524
19	Kalimantan Selatan	0.3917	0.9873	0.3069	0.2305	0.1881	0.5141
20	Kalimantan Timur	0.4177	0.9011	0.5867	0.6044	0.6106	0.3246
21	Kalimantan Utara	0.4833	0.9086	0.3751	0.4050	0.6217	0.5746
22	Sulawesi Utara	0.5420	0.8323	0.2236	0.5421	0.8026	0.1815
23	Gorontalo	0.4214	0.4176	0.1575	0.1121	0.8625	0.7298
24	Sulawesi Tengah	0.2598	0.5813	0.1405	0.3115	0.5657	0.7681
25	Sulawesi Barat	0.0106	0.6003	0.0735	0.0872	0.2886	0.8206
26	Sulawesi Selatan	0.2964	0.7704	0.1689	0.2928	0.7225	0.5161
27	Sulawesi Tenggara	0.2781	0.6050	0.1382	0.4548	0.8331	0.7379
28	Bali	1.0000	1.0000	0.4932	0.4953	0.9411	1.0000
29	NTB	0.0949	0.5309	0.1657	0.0685	0.8625	0.8105
30	NTT	0.0000	0.1197	0.0000	0.0623	0.5905	0.7520
31	Maluku	0.3368	0.3071	0.1659	0.6729	0.7617	0.1290
32	Maluku Utara	0.3368	0.8710	0.2592	0.4424	0.6037	0.5484
33	Papua	0.0257	0.1735	0.3316	0.7539	0.4932	0.0544
34	Papua Barat	0.1938	0.0000	0.3671	0.4953	0.3690	0.5282

Table 2 shows the normalized welfare indicators after applying Benefit–Cost transformation. The results indicate substantial variation across provinces, highlighting the heterogeneity of welfare conditions. This transformed dataset provides a standardized basis for the clustering process.

3.1.3 Determination of the Optimal Number of Clusters Using the Elbow Method

The optimal number of clusters was determined using the Elbow Method by analyzing the relationship between the number of clusters (k) and the within-cluster sum of squares (WCSS). The Elbow graph shows a significant decrease in WCSS up to k=4, after which the curve begins to flatten, forming an elbow point at this value. In addition, evaluation using the Silhouette Score yields a value of 0.2257 at k=4 indicating a reasonably good level of cluster separation. Based on these two approaches, this study determines four clusters as the optimal number.
The visualization of the Elbow Method is presented in Figure 2.

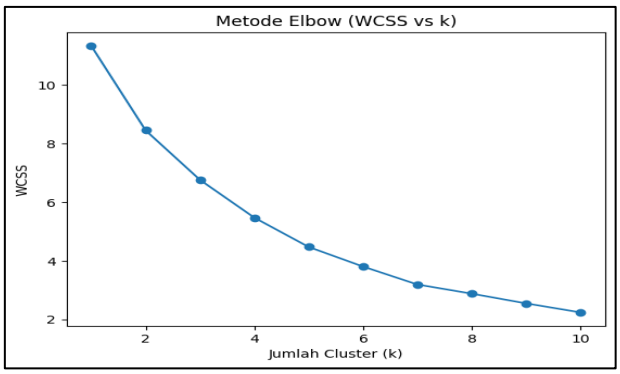


Figure 2. Elbow Method Results

The curve exhibits a clear elbow at k=4, indicating the point at which additional clusters no longer significantly reduce WCSS, thus confirming the optimal cluster number.

3.1.4 Clustering Using K-Means

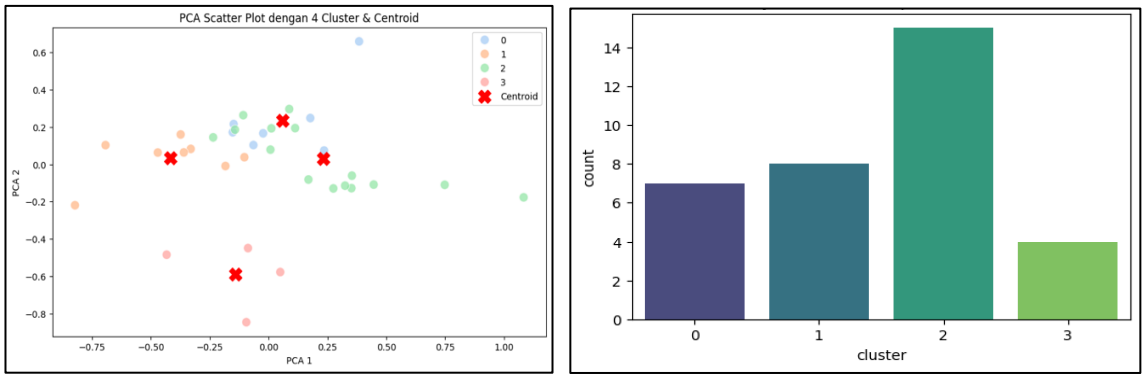
The clustering process in this study was conducted using the K-Means algorithm with a total of four clusters (K = 4), as determined in the previous optimal cluster selection stage. The dataset consists of provincial welfare indicators that have been normalized and adjusted to ensure consistent indicator direction, allowing each variable to contribute proportionally to the clustering process. The K-Means algorithm operates iteratively by initializing cluster centroids and subsequently updating their positions based on the mean of cluster members until convergence is achieved. In this study, convergence was reached after several iterations, indicated by stable centroid positions with no significant changes.

The clustering results are visualized using Principal Component Analysis (PCA) reduced into two dimensions, as shown in Figure 3(a). Each point represents a province, while different colors indicate cluster membership. The red cross markers represent the centroids of each cluster. From the visualization, it can be observed that:

1. Cluster 2 contains the largest number of members and exhibits a relatively wide distribution.
2. Cluster 3 has the smallest number of members and appears more separated from other clusters.
3. Clusters 0 and 1 show relatively balanced group sizes and are positioned closer to each other.

The distribution of provinces across clusters is presented in Figure 3(b). The distribution of provinces across clusters shows a noticeable imbalance, where certain clusters contain a larger number of provinces while others include fewer members. This indicates that some welfare

patterns are more dominant, whereas others represent more specific or extreme regional characteristics. Larger clusters reflect more common welfare conditions, while smaller clusters indicate more unique or localized patterns.



a **b**

Figure 3. K-Means Clustering Results: (a) PCA-based cluster visualization, (b) Distribution of provinces per cluster

Figure 3 presents the results of the K-Means clustering, including the PCA-based visualization in Figure 3(a) and the distribution of provinces across clusters in Figure 3(b). The visualization illustrates differences in cluster density and positioning, indicating variation in welfare characteristics among provinces. These results confirm the presence of diverse welfare patterns across regions. Table 3 presents the detailed grouping of provinces based on the clustering results.

Table 3. Results of Provincial Clustering Using the K-Means Algorithm

No	province	Clustering (K-Means)	No	province	Clustering (K-Means)
1	Aceh	3	5	Kep. Riau	0
2	Sumatera Utara	0	6	Jambi	2
3	Sumatera Barat	2	7	Sumatera Selatan	2
4	Riau	2	8	Kep. Bangka Belitung	2
9	Bengkulu	2	22	Sulawesi Utara	0
10	Lampung	2	23	Gorontalo	1
11	DKI Jakarta	0	24	Sulawesi Tengah	1
12	Jawa Barat	0	25	Sulawesi Barat	1
13	Jawa Tengah	1	26	Sulawesi Selatan	1
14	DIY	0	27	Sulawesi Tenggara	1
15	Jawa Timur	1	28	Bali	0
16	Banten	0	29	NTB	1
17	Kalimantan Barat	2	30	NTT	1
18	Kalimantan Tengah	2	31	Maluku	3
19	Kalimantan Selatan	2	32	Maluku Utara	2
20	Kalimantan Timur	0	33	Papua	3
21	Kalimantan Utara	2	34	Papua Barat	3

The interpretation of the characteristics of each cluster is discussed in the following section, focusing on the analysis of welfare indicator patterns within each group.

3.1 Discussion

This section interprets the clustering results and discusses their implications for regional welfare analysis.

3.2.1 Interpretation of Social Welfare Cluster Profiles

The clustering results divided the 34 provinces into four social welfare clusters based on six key indicators. Differences in indicator achievement patterns within each cluster reflect variations in welfare levels across regions. The average welfare indicator profiles within each cluster are presented in Table 4 as a basis for interpretation.

Table 4. Average Profile of Welfare Indicators in Each Cluster

Cluster	Stunting Prevention	Economic Welfare	Per Capita Expenditure	Average Years of Schooling	Access to Safe Drinking Water	Employment Engagement Index	Mean
0	0.6402	0.8644	0.4841	0.5860	0.8232	0.2881	0.6143
1	0.2758	0.5509	0.1352	0.1751	0.7171	0.6720	0.4210
2	0.4999	0.8115	0.2806	0.2861	0.4169	0.5087	0.4673
3	0.1765	0.2423	0.2559	0.6106	0.5673	0.2283	0.3468

The clustering results reveal four distinct groups of provinces with different welfare characteristics. Cluster 0 represents regions with a relatively high level of welfare, characterized by strong performance in economic, educational, and infrastructure dimensions. In contrast, Cluster 2 reflects regions with a moderate level of welfare, where economic and employment conditions are relatively strong, but limitations remain in education and basic infrastructure. Cluster 1 indicates regions with lower-middle welfare conditions, characterized by relatively low per capita expenditure and educational attainment despite higher workforce participation, suggesting the need to improve productivity and workforce quality. Meanwhile, Cluster 3 represents regions with the lowest level of welfare, facing significant challenges particularly in economic and health dimensions. These differences highlight the importance of a region-based development approach, where policy interventions should be tailored to the specific characteristics of each cluster. Regions in higher welfare clusters should focus on sustaining development, while lower welfare clusters require priority interventions in poverty reduction, public health improvement, and access to basic services. Furthermore, these findings are consistent with previous studies highlighting regional disparities in welfare across Indonesia, particularly in terms of access to education, infrastructure, and economic opportunities. However, this study extends prior research by integrating multidimensional indicators within a unified machine learning-based clustering framework, enabling a more comprehensive, objective, and data-driven classification of regional welfare compared to conventional single-indicator approaches.

Although the clustering results provide meaningful insights into regional welfare patterns, the Silhouette Score of 0.2257 indicates a moderate level of cluster separation. This suggests that some provinces may share similar characteristics across clusters, resulting in partial overlap between groups. Such conditions are common in socioeconomic data, where regional characteristics tend to be continuous rather than strictly separable. Recent studies have also reported similar challenges in achieving high cluster separation when analyzing complex social and regional datasets using K-Means clustering. For example, study [15] explained that socioeconomic indicators often produce partially overlapping clusters because provinces may simultaneously share similar economic, educational, and infrastructure characteristics. Likewise, study [13] emphasized that welfare indicators generally form continuous patterns rather than strictly separated groups, which affects cluster compactness and separation in regional clustering analysis.

Furthermore, the inclusion of multiple indicators in this study including health, economy, education, infrastructure, and employment contributes to increased data variability and cluster complexity. Provinces may demonstrate strong performance in one dimension while remaining weak in another, resulting in reduced inter-cluster distance. Therefore, despite the moderate Silhouette Score, the clustering results still provide meaningful and interpretable welfare patterns

across provinces in Indonesia. This finding suggests that moderate Silhouette values may still be acceptable in exploratory social data analysis involving heterogeneous and multidimensional characteristics.

Therefore, these findings should be interpreted with consideration of the inherent limitations of the clustering method. Future research may explore alternative approaches such as hierarchical clustering or density-based methods (e.g., DBSCAN), as well as the integration of temporal data to improve clustering performance.

4. Conclusions

This study models multidimensional social welfare across 34 provinces in Indonesia using the K-Means algorithm based on six key indicators. The evaluation results, using the Elbow Method and a Silhouette Score of 0.2257, indicate that the optimal number of clusters is four. The clustering results reveal clear differences in welfare levels, where the high-welfare cluster achieves an average indicator value of 0.6143, while the lowest cluster records 0.3468. The remaining two clusters fall within moderate welfare levels, with average values of 0.4673 and 0.4210, reflecting disparities across welfare dimensions. These findings confirm that social welfare in Indonesia is heterogeneous and cannot be adequately represented by a single indicator, highlighting the importance of a multidimensional and data-driven approach. From a practical perspective, the clustering results provide important implications for policymakers in designing targeted regional development strategies. Provinces classified in lower welfare clusters should be prioritized for interventions in poverty reduction, public health improvement, and expansion of access to education and basic infrastructure. Meanwhile, provinces in higher welfare clusters should focus on sustaining development performance while addressing internal inequality gaps.
For future research, it is recommended to explore more advanced clustering techniques, such as hierarchical or density-based methods, to better capture complex data structures. Additionally, integrating temporal and spatial data may enhance the accuracy and interpretability of welfare classification, enabling more dynamic and comprehensive policy analysis.

Acknowledgements

The author would like to thank Universitas Putra Indonesia (YPTK Padang) for the financial support provided through the SIMLITUPI 5 Grant Program for the 2025 Fiscal Year, under the Applied Research (PT) scheme. This support played a crucial role in supporting the implementation and completion of this research.

References

[1] I. Dinda Anjani And A. Bahtiar, "Penerapan Algoritma K-Means Clustering Untuk Mengelompokkan Penerima Bantuan Sosial Tunai (Bst) Di Jawa Barat," 2024. DOI : https://doi.org/10.36040/jati.v8i3.8974

[2] A. Az-Zahra And A. W. Wijayanto, "Tinjauan Kesejahteraan Di Daerah Perbatasan Republik Indonesia Tahun 2021: Penerapan Analisis Klaster K-Means Dan Hierarki," *Jurnal Sistem Dan Teknologi Informasi (Justin)*, Vol. 12, No. 1, P. 55, Jan. 2024, Doi: https://doi.org/10.26418/justin.v12i1.69040

[3] M. Bagus Herlambang And L. Theresia, "Pemetaan Kota/Kabupaten Endemis Demam Berdarah Dengue Dengan Analisis Data Science Menggunakan Algoritma Clustering," *Multimedia & Jaringan*, Vol. 8, No. 1, 2023. http://dx.doi.org/10.30811/jim.v8i1.3938

[4] H. H. Setiawan, J. Timur, K. Sosial, and P. Sosial, "Merumuskan indeks kesejahteraan sosial (iks) di indonesia defining social welfare index (swi) in indonesia," vol. 5, no. 200.DOI : https://doi.org/10.33007/inf.v5i3.1786

[5] A. Gafur, S. H. Rahmi, And F. Ahmadi, "Analisis Kesenjangan Ekonomi Daerah Perkotaan Dan Pedesaan Di Indonesia," *Economica Insight*, Vol. 2, No. 1, Pp. 31–42, Nov. 2025, Doi: https://doi.org/10.71094/ecoin.v2i1.203 .

[6] D. I. Yunistya, R. Goejantoro, F. Deny, And T. Amijaya, "The Application Of K-Harmonic Means Method In District/City Grouping (Case Study: Poverty In Kalimantan Island In 2020) Penerapan

Metode K-Harmonic Means Dalam Pengelompokan Kabupaten/Kota (Studi Kasus: Kemiskinan Di Pulau Kalimantan Tahun 2020)," Vol. 19, No. 1, Pp. 51–64, 2022, https://doi.prg/10.20956/J.V19i1.21116.

[7] N. Burkart And M. F. Huber, "A Survey On The Explainability Of Supervised Machine Learning," 2021.DOI : https://doi.org/10.1613/jair.1.12228 .

[8] P. Algoritma K-Means Dalam Pengelompokan Jumlah Penduduk Berdasarkan Kelurahan Di Kota Pematangsiantar, "Widodo Saputra 5) 1)2)3)4)5) Program Studi Teknik Informatika," Vol. 2, No. 2, Pp. 20–26, 2021, [Online]. DOI : https://doi.org/10.35960/ikomti.v2i2.704

[9] E. M. Y. A. G. D. W. P. D. S. W. Dan W. S. Dini Adni Navastara1), "Clustering Topik Penelitian Berbasis Unsupervised Learning Untuk Rekomendasi Koleksi Pustaka Di Perpustakaan Its," 2019.DOI : https://doi.org/10.12962/j24068535.v17i2.a788

[10] A. Zhu, Z. Hua, Y. Shi, Y. Tang, And L. Miao, "An Improved K-Means Algorithm Based On Evidence Distance," *Entropy*, Vol. 23, No. 11, Nov. 2021, Doi: https://doi.org/10.26418/bbimst.v12i3.67061

[11] H. Alexander, Y. Umaidah, And M. Jajuli, "Implementasi Clustering Untuk Menentukan Efektivitas Nilai Siswa Sesudah Pandemi Covid-19 Menggunakan Algoritma K-Means," 2023. DOI : https://doi.org/10.36040/jati.v7i3.7174

[12] R. A. Hakim, S. Harjanto, And A. Wibowo, "Penerapapan Metode K-Means Clustering Dalam Pengelolaan Kesejahteraan Sosial Di Desa Pentur," *Jurnal Teknologi Informasi Dan Komunikasi (Tikomsin)*, Vol. 12, No. 1, P. 32, Apr. 2024, Doi: 10.30646/Tikomsin.V12i1.820. DOI : http://dx.doi.org/10.30646/tikomsin.v12i1.820

[13] Y. L. Dakhi And B. A. Ningsi, "Pengelompokan Kabupaten Dan Kota Provinsi Sumatera Utara Berdasarkan Indikator Kesejahteraan Rakyat Menggunakan Algoritma K-Means," *Malcom: Indonesian Journal Of Machine Learning And Computer Science*, Vol. 4, No. 3, Pp. 993–1003, Jun. 2024, Doi: https://doi.org/10.57152/malcom.v4i3.1381

[14] Y. Sihombing, "Inovasi Kelembagaan Pertanian Dalam Mewujudkan Ketahanan Pangan," *Proceedings Series On Physical & Formal Sciences*, Vol. 5, Pp. 83–90, 2023, Doi: https://doi.org/10.30595/pspfs.v5i.707

[15] Y. Febby Rachmawati, "Analisis Cluster Menggunakan Metode K-Means Dan Fuzzy C-Means Pada Faktor Sosial Ekonomi Di Kalimantan Barat," 2024. DOI : https://doi.org/10.26418/bbimst.v13i5.81858

[16] P. Algoritma K-Means Dalam Pengelompokan Kesejahteraan Rakyat Di Kabupaten Karawang *Et Al.*, "Penerapan Algoritma K-Means Dalam Pengelompokan Kesejahteraan Rakyat Di Kabupaten Karawang". DOI : https://doi.org/10.35889/progresif.v17i2.649

[17] M. Ihza Zuhendra And R. Hidayat, "Penerapan Data Mining Untuk Klasterisasi Tingkat Kemiskinan Berdasarkan Data Terpadu Kesejahteraan Sosial (Dtks)," Vol. 7, No. 1, Pp. 32–41, 2024. DOI : https://doi.org/10.36080/skanika.v7i1.3149

[18] Kristanto Setyo Utomo, "Perbandingan Algoritma Machine Learning Untuk Penentuan Klasifikasi Kemiskinan Multidimensi Di Provinsi Nusa Tenggara Timur," 2022, Doi: 10.5300/Jstar.V2i01.24. DOI : https://doi.org/10.5300/JSTAR.V2I01.24

[19] I. Yati Beti And H. Juliansa, "Klik: Kajian Ilmiah Informatika Dan Komputer Penerapan Normalisasi Data Metode Decimal Scaling Dan Metode K-Means Dalam Mengelompokkan Kasus Demam Berdarah," *Media Online)*, Vol. 4, No. 6, Pp. 2928–2936, 2024, Doi: https://doi.org/10.30865/klik.v4i6.1925

[20] A. L. Firmansyah, B. I. Nugroho, And Z. Arif, "Optimasi K-Means Clustering Pada Data Harga Mangga Menggunakan Particle Swarm Optimization," *Jurnal Teknologi Sistem Informasi*, Vol. 6, No. 2, Pp. 245–259, Sep. 2025, Doi: http:///doi.org/10.35957/Jtsi.V6i2.13158