UJM 13 (2) 2024
UNNES Journal of Mathematics
http://journal.unnes.ac.id/sju/index.php/ujm

A Data Mining Approach to Wage Inequality Analysis in Indonesia: A Clustering Study Using Fuzzy C-Means

Nur Achmey Selgi Harwanti*, Putriaji Hendikawati, Ratna Nur Mustika Sanusi, Alfian Adi Pratama

Department of Mathematics, Faculty of Mathematics and Natural Science,
Universitas Negeri Semarang, Semarang, 50299, Indonesia

Article Info

Article History:
Recived …….
Accepted …..
Published …..

Keywords:
Fuzzy C-Means Clustering, Regional Economy, Wage Analysis

Abstract

This study aims to cluster Indonesian provinces based on the average wage structure of workers across 17 economic sectors using the Fuzzy C-Means (FCM) method. The wage data underwent preprocessing steps including missing value imputation using the median, logarithmic transformation to reduce skewness, and Z-Score standardization to ensure uniform data scaling. The evaluation of the number of clusters and fuzziness values was conducted using the Silhouette coefficient and Fuzzy Partition Coefficient (FPC), with the best results achieved at three clusters and a fuzziness value of 1.3. Further analysis using Principal Component Analysis (PCA) provided visualization of the clusters, while radar charts illustrated wage characteristics by sector within each cluster. The clustering results reveal significant economic disparities among provinces: Cluster 1 consists of provinces with the highest wages dominated by high-value-added sectors such as mining and finance; Cluster 0 shows a balanced wage distribution reflecting a transitional economy; and Cluster 2 includes provinces with the lowest wages facing structural challenges. These findings offer a comprehensive overview of regional economic diversity in Indonesia and can serve as a basis for policy-making aimed at more equitable economic development.
.

How to cite:

Harwanti, N. A. S., Hendikawati, P., Sanusi, R. N. M., & Pratama, A. A. 2024. A Data Mining Approach to Wage Inequality Analysis in Indonesia: A Clustering Study Using Fuzzy C-Means. *UNNES Journal Of Mathematics*. 13(2): 39-47.

* Sekaran, Gunung Pati, Jawa Tengah, Indonesia
E-mail: selgi@mail.unnes.ac.id

p-ISSN 2252-6943
e-ISSN 2460-5859

1. Introduction

Achieving equitable welfare remains one of the central challenges in Indonesia's national development agenda. Equity in welfare not only pertains to income distribution but also encompasses access to employment opportunities, quality education, healthcare services, and economic infrastructure. One of the key indicators for assessing welfare levels is labor wage. Wages serve as compensation for workers' contributions to the production process and play a vital role in determining purchasing power, household consumption, and overall individual or family well-being. Therefore, examining wage distribution and structure is essential in evaluating the effectiveness of economic development policies at both national and regional levels.

In Indonesia, wage levels vary significantly across provinces and economic sectors, reflecting differences in labor productivity, the degree of industrialization, as well as disparities in access to education and infrastructure. Provinces such as DKI Jakarta, West Java, and East Kalimantan—characterized by advanced industrial and service sectors—tend to have higher wage levels compared to regions dominated by agriculture or the informal economy. Research by (Winardi, Proyarsono, Siregar, & Kustanto, 2021) that the development of industrial estates, such as those in West Java, has significantly contributed to employment generation not only within the industrial sector itself but also across other production sectors, ultimately reinforcing disparities in regional welfare.

Data from the Central Statistics Agency (BPS) show that the average wages of workers in Indonesia vary significantly across provinces and economic sectors. Sectors such as agriculture, forestry, and fisheries; manufacturing; construction; financial services and insurance; education; as well as transportation and warehousing each exhibit distinct wage characteristics. These disparities are influenced by several factors, including productivity levels, business scale, technology adoption, market demand, and labor regulations. For instance, the financial services sector typically offers higher wages due to the requirement for advanced skills and higher educational attainment, whereas the agricultural sector tends to have lower wages as it remains dominated by informal and traditional labor practices.

Wage data by economic sector across provinces can provide a comprehensive overview of the regional economic structure and the potential disparities in welfare that may arise. While previous studies have focused on clustering based on education indicators, minimum wages, or quality of life indexes, this study introduces a novel approach by specifically clustering provinces based on the average wages across 17 economic sectors, enabling a more detailed and multi-sectoral understanding of regional economic typologies. This modification allows the analysis to capture intra-regional disparities in sectoral wages, which are often overlooked in studies focusing on aggregated or macro indicators alone.

Thus, the concrete problem addressed in this research is how to classify Indonesian provinces into meaningful clusters based on similarities in wage structures across diverse economic sectors, in order to reveal hidden regional economic patterns and disparities that are not captured through conventional approaches. The study also aims to determine the optimal number of clusters and the appropriate fuzziness level for this classification using internal validation indices. Analyzing this data is crucial for development planning that prioritizes social equity. One statistical approach that can be employed to identify such patterns is cluster analysis.

Cluster analysis is an exploratory method in multivariate statistics aimed at grouping objects into several clusters based on the degree of similarity among them. Unlike classification methods, which require predefined class labels, cluster analysis is a form of unsupervised learning that does not rely on prior information about group membership. This technique is useful for uncovering hidden structures within data and has been widely applied across various fields, including economics, marketing, and public health.

In general, cluster analysis methods can be categorized into two main groups: hierarchical and non-hierarchical methods. Hierarchical methods form clusters gradually through either agglomerative (merging) or divisive (splitting) processes, whereas non-hierarchical methods, such as K-Means and Fuzzy C-Means (FCM), create clusters by centering data points around specific centroids. FCM is a non-hierarchical variant that excels in handling data with ambiguous characteristics or overlapping clusters (Askari, 2021).

Fuzzy C-Means, first introduced by Bezdek (1981), is a clustering technique based on fuzzy logic theory. Unlike traditional clustering methods where each object belongs exclusively to one cluster, FCM allows objects to have membership degrees across multiple clusters. These membership values range between 0 and 1, indicating the likelihood that an object belongs to each cluster. This approach is particularly suitable for social and economic data, where boundaries between groups are often indistinct (Bezdek, 1981).

The Fuzzy C-Means algorithm operates by minimizing the following objective function:

$$\boldsymbol{J(X,U,V)} = \sum_{k=1}^{n} \sum_{i=1}^{c} (\boldsymbol{u_{ik}})^{m} \boldsymbol{D_{ik}^{2}} \qquad \textbf{(1)}$$

In each iteration of the algorithm, the cluster centers and membership degrees of the objects are updated until the objective function reaches convergence. The strength of this method lies in its ability to capture uncertainty in the clustering process, which more accurately reflects real-world conditions in socio-economic data.

Once the clustering process is completed, evaluating the quality of the resulting clusters becomes essential. One commonly used method for cluster validation is the Silhouette Coefficient, which measures how well an object fits within its assigned cluster compared to other clusters (Lai, Huang, & Lu, 2025). A high Silhouette value indicates that an object is appropriately grouped with its cluster and is relatively distant from others, making it a reliable indicator of clustering validity.

The application of this method can produce a classification of regions based on similarities in wage characteristics and industrial sectors, thereby providing valuable insights for regional development planning. Several studies have employed clustering techniques to identify patterns of regional inequality in Indonesia. For instance, (Nasrullah, Widodo, & Syahputra, 2023) evaluated the use of the Fuzzy C-Means method to cluster regions in South Sulawesi Province based on education indicators and minimum wage levels. Their findings revealed that the regions could be grouped into two main

clusters, offering valuable input for policy formulation in education resource allocation. Similarly, a study by (Suci, Fadillah, & Ramadhan, 2023) employed hierarchical clustering with a complete linkage technique to classify districts and municipalities in Jambi Province based on education levels and average wages by main occupational sectors. The results identified three distinct clusters with varying characteristics in terms of education and wage levels, providing a detailed picture of intra-provincial development disparities.

In addition, (Rahman, Pratiwi, & Nugraha, 2023) examined the clustering of districts and municipalities in Central Java Province using both K-Means and Fuzzy C-Means methods. The results indicated that the K-Means method yielded the best SW and SB ratio; however, the Fuzzy C-Means method also proved effective in mapping regions based on similarities in economic characteristics. These findings highlight the importance of selecting an appropriate clustering method to ensure accurate outcomes. Another study by (Amalia, Hidayat, & Wulandari, 2023) applied K-Means clustering to classify districts and municipalities in West Java Province based on indicators of quality of life, such as population density, per capita expenditure, and regional minimum wages. The clustering results revealed three main groups, which can serve as a foundation for formulating regional development policies that are more responsive to local conditions. Beyond clustering approaches, (Badaruddin & Saadah, 2022) emphasized that wage disparities in Indonesia are also closely linked to labor productivity across various sectors and regions. These findings underscore the need for a more comprehensive regional analysis to support equitable and inclusive development policies.

Furthermore, a study by (Miranti & Mendez, 2022) revealed significant social and economic convergence at the district and municipal levels in Indonesia during the period 2010–2018. Using a spatial panel data approach, the study found that a region's Human Development Index (HDI) and Gross Regional Domestic Product (GRDP) per capita tended to increase more rapidly if neighboring regions exhibited high HDI and GRDP levels. These findings reinforce the importance of accounting for spatial effects when analyzing regional disparities, including in the context of sectoral wage distribution. Their results also indicated that social convergence occurred slightly faster than economic convergence, with the industrial and service sectors contributing significantly to social convergence, while initial economic size influenced economic convergence.

Given the substantial variation in wage conditions across sectors and regions in Indonesia, as well as the potential spatial interdependence among provinces or districts, this study distinguishes itself from prior works by focusing on a broader and more sectorally disaggregated wage dataset (17 economic sectors), and by employing Fuzzy C-Means clustering with internal validation techniques (Silhouette Coefficient and Fuzzy Partition Coefficient) to identify optimal cluster configurations. This approach is expected to support the formulation of more equitable, adaptive, and data-driven development policies.

2. Method

This study adopts a quantitative approach using an exploratory method through cluster analysis to group provinces in Indonesia based on the average wage of workers by industry sector. The data used in this study are secondary data obtained from the official publication of Statistics Indonesia (Badan Pusat Statistik/BPS) in 2024. The dataset includes the average wages of workers in 38 provinces, categorized by several main industry sectors with the following variables:

Table 1. Research Variables

Variable	Keterangan
X1	Agriculture, Forestry, and Fisheries
X2	Mining and Quarrying
X3	Manufacturing Industry
X4	Electricity and Gas Supply
X5	Water Supply, Waste Management, Waste Disposal, and Recycling
X6	Construction
X7	Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles
X8	Transportation and Warehousing
X9	Accommodation and Food Service Activities
X10	Information and Communication
X11	Financial and Insurance Activities
X12	Real Estate Activities
X13	Professional, Scientific, and Technical Activities
X14	Public Administration, Defense, and Compulsory Social Security
X15	Education Services
X16	Human Health and Social Work Activities
X17	Other Services

The data analysis technique employed in this study is Fuzzy C-Means (FCM) Clustering, a non-hierarchical clustering method. Unlike the K-Means algorithm, which assigns each object exclusively to a single cluster, FCM allows each object to have degrees of membership across multiple clusters, with values ranging from 0 to 1. This method is considered particularly relevant for socio-economic data, as it can effectively capture uncertainty and overlapping characteristics among groups, which frequently occur in real-world contexts.

The analytical steps in this study were carried out systematically through the following stages:

a. Data Collection

Secondary data were collected from official publications of Statistics Indonesia (Badan Pusat Statistik/BPS), containing information on the average wages of workers by province and economic sector for the year 2023 (BPS, 2023).

b. Preprocessing Data
   - Checking the completeness and consistency of the data.
   - Removing irrelevant or duplicate entries, if any.
   - Normalizing the data to prevent bias due to differences in variable scales, using either min-max scaling or z-score normalization.

c. Determination of Initial Number of Clusters
   The initial number of clusters (value of c) was determined as an input for the Fuzzy C-Means algorithm. This step was conducted in an exploratory manner by considering domain knowledge and evaluating cluster validity indices.

d. Implementation of Fuzzy C-Means Clustering
   The clustering process was performed using the Fuzzy C-Means algorithm in R software. The key parameters included:
   - Number of clusters (c)
   - Fuzzification parameter (m), usually set between 1.5 and 2.5
   - Convergence tolerance threshold and the maximum number of iterations

e. Evaluation of Cluster Quality
   The quality of the clustering results was assessed using the Silhouette Coefficient, which provides information on the compactness and separation of clusters. This value also assists in identifying the optimal number of clusters.

f. Interpretation and Visualization of Clusters
   Each cluster was interpreted based on the dominant wage characteristics in each economic sector, and cluster visualizations were created to facilitate understanding.

g. Conclusion and Policy Recommendation
   The clustering results were linked to regional socio-economic conditions, and cluster-based policy recommendations were formulated, such as focusing on wage improvement in specific sectors or developing regionally differentiated wage policies.

3. Results and discussions

3.1. Data Preprocessing and Charateristics

Prior to conducting cluster analysis, the dataset on labor wages by province underwent a preprocessing phase to ensure data quality and completeness. One of the key steps in this process was identifying and handling missing values, as these can negatively impact the validity of the analysis. A data check revealed the presence of missing values in several variables, which required appropriate treatment before further analysis could proceed. Specifically, missing values were found in variables X2, X3, X5, X7, X9, and X12. To address this issue, the missing data in these variables were imputed using the median. This approach was chosen to provide a more robust estimate that is less affected by potential outliers, thereby maintaining the overall reliability of the dataset.

After the imputation process, descriptive statistical exploration was conducted to illustrate the characteristics of labor wages across sectors in all provinces. The statistical measures used included the mean, standard deviation, median, minimum, and maximum values. The results are presented in Table 2 below.

Table 2. Descriptive Statistics of Average Labor Wages per Sector (in IDR)

Variable	Mean	Standard Deviation	Median	Minimum	Maximum
X1	2.582.506,84	606.614,56	2.596.742,09	1.421.476,04	4.034.623,50
X2	5.126.241,28	2.499.768,98	4.275.008,17	1.474.264,05	11.871.502,55
X3	3.030.527,40	1.006.041,91	2.948.228,52	1.391.865,72	5.853.700,30
X4	4.573.167,45	1.343.122,63	4.324.645,08	2.614.762,83	8.921.110,97
X5	3.079.216,95	1.094.803,69	2.899.772,96	1.500.484,50	6.406.143,19
X6	3.243.769,24	866.191,63	3.050.581,71	2.043.381,08	5.847.647,75
X7	2.722.270,91	689.370,70	2.616.549,21	1.767.440,03	5.357.278,48
X8	3.567.329,48	1.160.764,59	3.333.599,81	1.824.593,56	7.409.324,27
X9	2.357.644,89	796.883,84	2.071.649,32	1.286.564,81	4.808.109,35
X10	3.704.115,24	1.512.664,01	3.299.101,17	1.986.561,06	8.175.539,50
X11	4.460.953,45	1.282.987,30	4.325.478,79	2.689.000,00	9.957.484,84
X12	3.761.288,14	786.045,55	3.761.288,14	1.703.904,56	5.830.219,50
X13	3.330.820,25	1.074.858,37	3.020.016,08	1.924.062,08	7.579.479,84
X14	4.069.126,25	874.713,65	3.771.573,79	3.092.639,84	7.012.423,83
X15	3.196.831,67	606.379,79	3.105.565,46	2.237.322,76	4.958.573,41
X16	3.614.553,83	1.063.851,78	3.331.278,95	2.169.446,10	7.444.175,17
X17	2.039.407,39	641.704,25	1.867.316,29	1.204.473,41	4.154.566,54

The data distribution was also analyzed using skewness values. For variables with |skewness| ≥ 1, a transformation was applied using the natural logarithm function based on the log1p method (i.e $x_{transform} = \log(1 + x)$ to reduce skewness.

Following the transformation, all numerical variables were standardized using Z-Score Standardization to ensure consistency in scale across variables. This process converts each value into a z-score based on the mean and standard deviation of the variable, resulting in variables with a mean of 0 and a standard deviation of 1. The standardization formula used is as follows:

$$\boldsymbol{z = \frac{x - \mu}{\sigma}}$$

Where $\boldsymbol{x}$ is the original value, $\boldsymbol{\mu}$ is the mean, and $\boldsymbol{\sigma}$ is the standard deviation of the variable. The outcome of this process is a dataset in a uniform scale, suitable for cluster analysis. The results of the transformation process are presented below.

The results of the transformation process are visualized using histograms for several variables identified as having skewed distributions. These histograms illustrate a comparison between the data distribution before and after the application of logarithmic transformation and standardization.

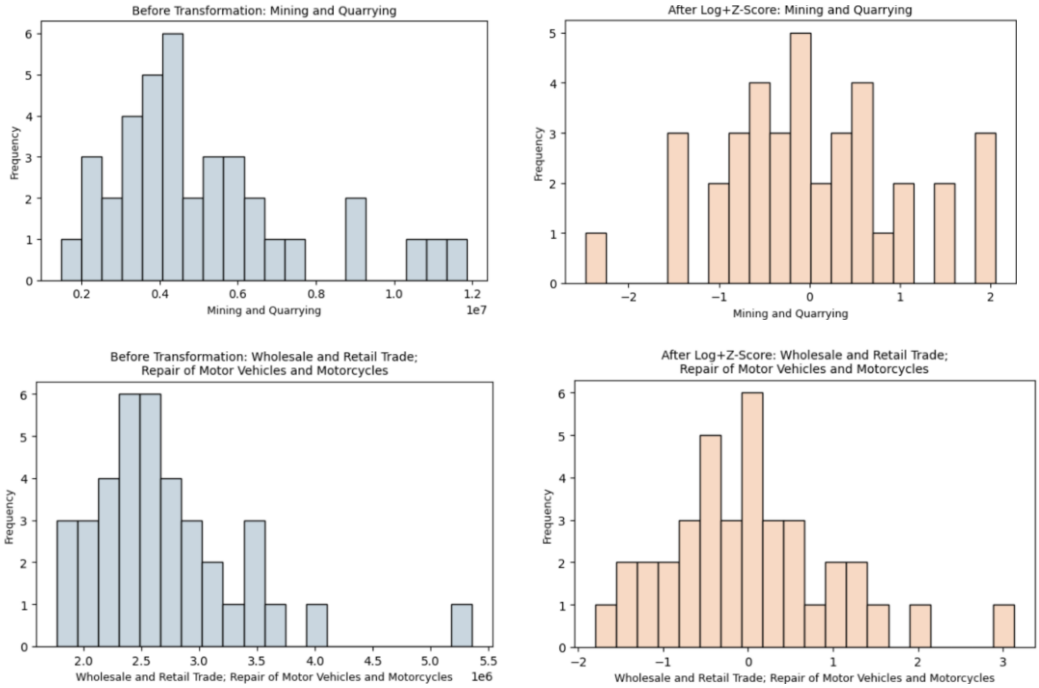

Figure 1. comparison between the data distribution before and after the application of logarithmic transformation and standardization**.**

3.2. Performance Evaluation Based on the Number of Clusters

In this stage, a clustering process was conducted for 38 provinces in Indonesia based on the average hourly wages across 17 categories of economic sectors. The objective of this process was to identify provinces with similar sectoral wage structures, allowing for an analysis of patterns of economic similarity or divergence across regions.

The clustering was performed using the Fuzzy C-Means (FCM) method. Unlike hard clustering methods such as K-Means, FCM allows each province to have degrees of membership in multiple clusters. This provides a more flexible and realistic representation of the heterogeneous structure among regions. Moreover, the fuzziness coefficient (m) influences the clustering outcome, where a lower fuzziness value results in more distinct memberships, while a higher value leads to more diffuse memberships across clusters.

In this analysis, the number of clusters was varied from 3 to 6, with different fuzziness values set at 1.3, 1.5, 1.7, and 2.0. The purpose of using a range of cluster numbers was to explore the optimal grouping based on the Silhouette Coefficient, which measures how well each province fits within its assigned cluster. The results of the analysis based on the number of clusters and fuzziness values are presented below.

Table 4. Evaluation of the Number of Clusters and Fuzziness

Cluster	Fuzziness	FPC	Silhouette
3	1,3	0,8762	0,2247
4	1,3	0,7992	0,1390
5	1,3	0,8001	0,1352
6	1,3	0,7494	0,1089
3	1,5	0,7366	0,2247

4	1,5	0,6033	0,1349
5	1,5	0,5787	0,1352
6	1,5	0,4876	0.0936
3	1,7	0,6141	0,2247
4	1,7	0,4757	0,1349
5	1,7	0,3816	0,0971
6	1,7	0,3564	0,0785
3	2,0	0,4941	0,2247
4	2,0	0,3715	0,1390
5	2,0	0,2943	0,0895
6	2,0	0,2548	0,0589

From the analysis results presented in the table above, it can be observed that using 3 clusters with a fuzziness value of 1.3 yields a relatively good Silhouette Coefficient of 0.225, along with a relatively high Fuzzy Partition Coefficient (FPC) of 0.8762. This indicates that the provinces within the three clusters exhibit a fairly strong similarity in terms of sectoral wage structures compared to solutions with a greater number of clusters.

However, when using a fuzziness value of 1.5, the 3-cluster solution still achieves the same high Silhouette Coefficient of 0.225 and a reasonably high FPC of 0.735, suggesting that the cluster quality remains good with clear separation between clusters. At fuzziness values of 1.7 and 2.0, although the FPC values decrease somewhat, the Silhouette Coefficient remains consistent at approximately 0.2247. The 3-cluster solutions with fuzziness values of 1.7 and 2.0 still show similarities with those at lower fuzziness levels, albeit with a slight reduction in cluster separation quality as indicated by the lower FPC scores.

Based on this evaluation, the 3-cluster solution with a fuzziness value of 1.3 is considered the best choice, as it balances cluster separation quality and the similarity among provinces within the formed clusters.

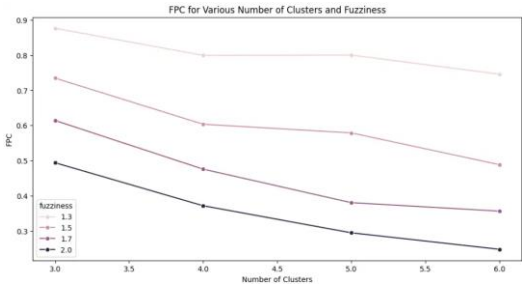

Figure 2. FPC Graph for Various Numbers of Clusters and Fuzziness Values

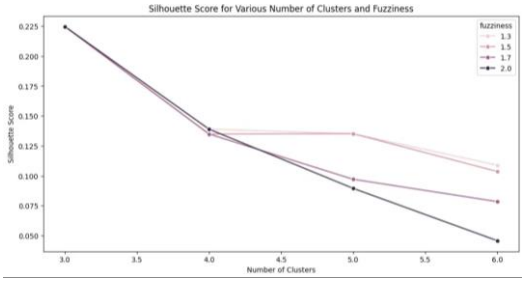

Figure 3. Silhouette Score for Various Numbers of Clusters and Fuzziness Values

3.3. Clustering Results

Based on the cluster analysis using the Fuzzy C-Means (FCM) method with three clusters and a fuzziness value of 1.3, the provinces in Indonesia were grouped according to similarities in their sectoral wage structures. To facilitate the interpretation of the clustering results, dimensionality reduction was performed using Principal Component Analysis (PCA), allowing the data to be visualized in two- and three-dimensional spaces.

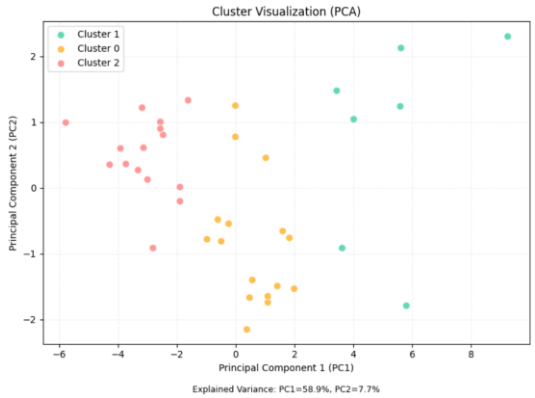

Figure 4. 2D PCA Graph

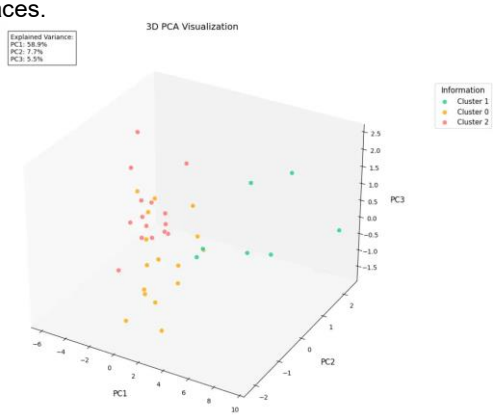

Figure 5. 3D PCA Graph

The visualization of the dimensionality reduction results using PCA is presented in Figures 4 and 5. Each point represents a province, with different colors indicating the clusters formed through the clustering process. Based on the 3D visualization in Figure 7, the clusters tend to form distinct groups within the three-dimensional principal component space, although some slight overlap between clusters is observed. Figure 8, which shows the 2D projection of the first two principal

components (PC1 and PC2), demonstrates clearer cluster separation, particularly between cluster 2 (green) and the other clusters. This indicates that provinces within the same cluster share similar wage distribution patterns across the main economic sectors, while separated clusters suggest significantly different wage distribution characteristics.

Furthermore, to identify the characteristics of each cluster, an analysis of the average wages per sector for each cluster was conducted. The results of this analysis are visualized in the form of radar charts in Figure 6.

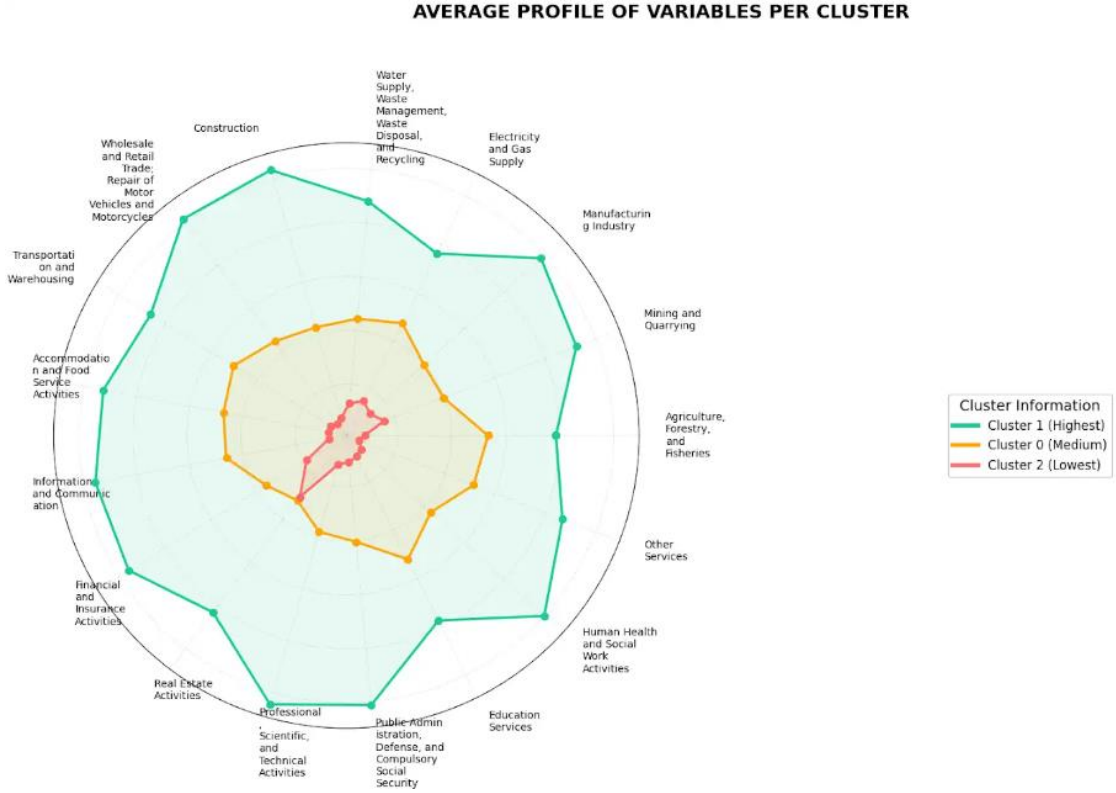

Figure 6. Radar Chart of Average Wages per Sector in Each Cluster

The radar chart illustrating the average wages per sector across each cluster reveals significant economic disparities among Indonesian provinces. Cluster 1 comprises provinces with the highest wage levels in nearly all sectors, indicating regions with more advanced and diversified economic structures dominated by high value-added sectors such as mining, finance, and information and communication. For example, DKI Jakarta and West Java, which belong to this cluster, have rapidly growing service and industrial sectors that require skilled labor compensated with relatively high wages. Additionally, East Kalimantan and Papua, rich in natural resources, offer high wages in the mining sector due to the high productivity and profitability of extractive activities.

From the average wage data, it is evident that the mining and finance sectors in Cluster 1 pay substantially higher wages compared to other clusters. For instance, the average mining sector wage in Cluster 1 is approximately 8.6 million rupiahs, whereas it is about 4.9 million and 3.6 million rupiahs in Clusters 0 and 2, respectively. This wage gap reflects the reliance of these regions on natural resources and capital-intensive industries that drive higher labor income. However, this dependence also poses economic risks if global commodity prices fluctuate or investment declines.

Cluster 0 shows a more balanced wage distribution, generally positioned between the highest and lowest clusters. Provinces such as Bali and Yogyakarta in this cluster exhibit economic patterns combining tourism services with developing manufacturing and trade sectors. This suggests an economy in transition, where wages in traditional sectors are gradually rising alongside the growth of more productive formal and service sectors. Nonetheless, wages in sectors like accommodation and trade remain relatively low, indicating challenges in improving productivity and job quality in the informal and tourism sectors.

Cluster 2, with the lowest average wages across nearly all sectors, includes provinces facing structural challenges such as low investment, limited infrastructure, and dominance of informal or subsistence agriculture sectors. Provinces like Central Java, East Java, as well as parts of Nusa Tenggara and Sulawesi, show low wages in manufacturing, trade, and service sectors. This pattern reflects a lack of formal employment opportunities with decent wages and low technology and innovation penetration. These limitations also contribute to higher poverty rates and lower purchasing power in these regions.

To provide a clearer overview of the composition of each cluster, Table 6 presents the list of provinces included in Clusters 0, 1, and 2 along with the number of provinces in each cluster. This presentation helps to identify the spatial distribution of provinces based on similarities in sectoral wage structures. Such information is important as a basis for formulating more targeted development policies that align with the economic characteristics of each provincial group.

Table 5. Clustering Results for Each Province

Cluster	Number of Provinces	List of Provinces
Highest	7	Kep. Riau, DKI Jakarta, Jawa Barat, Banten, Kalimantan Timur, Papua, Papua Tengah.
Medium	16	Riau, Sumatera Selatan, Kep. Bangka Belitung, DI. Yogyakarta, Bali, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan, Maluku Utara, Papua Barat, Papua Barat Daya, Papua Selatan, Papua Pegunungan.
Lowest	15	Aceh, Sumatera Utara, Sumatera Barat, Jambi, Bengkulu, Lampung, Jawa Tengah, Jawa Timur, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku

Additionally, the clustering results are visualized geographically through a thematic map (Figure 7), facilitating the identification of spatial patterns of province groupings based on sectoral wage structures. This mapping serves as a strategic tool for regional development planning that takes into account economic disparities between regions.

Figure 7. Map of Indonesian Provinces Based on Clusters of Average Monthly Wages

Geographically, the highest wage cluster is concentrated mostly in western Java and resource-rich areas of Kalimantan and Papua. This demonstrates an economic centralization in regions with better access to capital, technology, and markets. Conversely, the lowest wage cluster is scattered across eastern Indonesia and much of the interior of Sumatra and Sulawesi, areas historically challenged by inadequate infrastructure and limited educational access.

The causes of this wage disparity can be traced to structural factors such as differing capacities of regions to attract investment and develop high value-added industries, quality of human resources, and local government policies. Regions that successfully develop formal and innovative sectors can raise labor wages and improve living standards, while others lag behind due to insufficient infrastructure and market access. This underscores the need for targeted development interventions focused on enhancing education quality, infrastructure development, and creating economic ecosystems that support productive sectors in lagging regions.

The novelty of this study lies in the use of Fuzzy C-Means clustering on 17 sectoral wage variables across 38 provinces—providing a more granular and multidimensional perspective of regional economic inequality in Indonesia. Unlike previous studies that used one or two indicators (e.g., minimum wage, education), this research captures a broader structure of labor income across various economic sectors, enabling a richer interpretation of regional similarities and disparities. The combination of internal cluster validation (FPC and Silhouette) and PCA visualization further enhances the robustness and interpretability of the clustering results.

Interpretively, the findings reveal that provinces cannot be simply grouped based on geography or development status alone. For instance, some provinces in eastern Indonesia, such as Papua Tengah, share wage structure similarities with economically advanced western provinces due to the influence of high-paying extractive industries. This finding challenges the assumption of linear east-west economic inequality and highlights the role of sectoral composition in shaping regional income patterns.

Furthermore, the cluster characterization adds actionable insight—Cluster 1 requires policies that sustain high-value sectors while managing volatility (e.g., global commodity price risks), Cluster 0 could benefit from investment and upskilling to accelerate sectoral transitions, and Cluster 2 needs foundational improvements in infrastructure, education, and access to formal employment.

In sum, the interpretation of the clustering outcomes supports the formulation of differentiated policy interventions that reflect each region's wage structure and economic composition. These results provide a more nuanced alternative to conventional development planning approaches that often rely on aggregate indicators alone.

4. Conclusion

The cluster analysis using the Fuzzy C-Means (FCM) method successfully identified patterns of economic disparity among Indonesian provinces by grouping them based on wage structures across 17 economic sectors. The FCM approach, which allows each province to belong to multiple clusters with varying degrees of membership, proves to be effective in capturing the complex and overlapping nature of regional economies in Indonesia.

The analysis produced three distinct clusters: Cluster 1 includes provinces with the highest average wages, dominated by advanced manufacturing and high-value services, indicating strong investment and high labor productivity. Cluster 2 reflects provinces with moderate and balanced wage levels, where a mix of formal and informal sectors coexist, often supported by local policy efforts toward economic diversification. Cluster 3 consists of provinces with the lowest average wages, typically characterized by informal labor dominance, low productivity, and limited access to infrastructure and capital.

These clustering results reveal a clear segmentation of Indonesia's labor market, highlighting the regional inequalities in wage distribution and sectoral development. The data preprocessing steps, such as imputation, log transformation, and standardization, enhanced the clustering accuracy by reducing data skewness and scale inconsistencies. In conclusion, the application of FCM has not only uncovered the structural wage differences among provinces but also provided a more nuanced understanding of transitional economic characteristics. These insights offer a valuable foundation for designing region-specific economic policies aimed at reducing wage gaps, promoting inclusive growth, and improving labor market equity throughout Indonesia.

.

References

Amalia, N., Hidayat, R., & Wulandari, R. (2023). nalisis K-Means Clustering pada Pengelompokan Kabupaten/Kota di Jawa Barat Berdasarkan Indikator Kenyamanan Hidup. *Jurnal Statistika dan Sains Data*, 35-44.

Askari, S. (2021). Fuzzy C-Means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: Review and development. *Expert Systems with Applications*.

Badaruddin, M., & Saadah, N. (2022). Analisis Ketimpangan Upah dan Produktivitas Tenaga Kerja di Indonesia. *Jurnal Ekonomi Pembangunan*, 101-115.

Bezdek, J. (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. *Springer Science & Business Media*.

BPS. (2023). *Statistik Upah Pekerja Indonesia 2023.* Jakarta: BPS RI.

Lai, H., Huang, T., & Lu, B. (2025). Silhouette coefficient-based weighting k-means algorithm. *Neural Comput & Applic 37*, 3061-3075.

Miranti, R., & Mendez, C. (2022). Social and Economic Convergence Across Districts in Indonesia: A Spatial Econometric Approach. *Bulletin of Indonesian Economic Studies*, 421-445.

Nasrullah, N., Widodo, H., & Syahputra, M. (2023). Klasterisasi Wilayah di Sulawesi Selatan Berdasarkan Indikator Pendidikan dan Upah Minimum Menggunakan Fuzzy C-Means. *Jurnal Sains dan Statistika*, 55-56.

Rahman, H., Pratiwi, Y., & Nugraha, M. (2023). Perbandingan Metode K-Means dan Fuzzy C-Means dalam Klasterisasi Kabupaten/Kota di Provinsi Jawa Tengah. *Jurnal Statistika Terapan dan Komputasi*, 72-80.

Suci, A., Fadillah, N., & Ramadhan, D. (2023). Hierarchical Clustering untuk Klasterisasi Kabupaten/Kota di Provinsi Jambi Berdasarkan Tingkat Pendidikan dan Upah Rata-rata. *Jurnal Matematika dan Aplikasinya*, 98-107.

Winardi, Proyarsono, D., Siregar, H., & Kustanto, H. (2021). The Impact on Industrial Estate Development Policy to Employment Absorption of West Java Province. *Jurnal Ekonomi Pembangunan*, 230-241.