**Multidimensional Poverty Determinants and Spatial Clustering of Districts in North Sumatra, Indonesia: A Factor and Cluster-Based Analytical Approach**

Abstract :

This study aims to identify the multidimensional determinants of poverty and classify regions in North Sumatra Province, Indonesia, based on shared socio-economic characteristics. The goal is to support targeted and region-specific poverty alleviation policies. A quantitative approach was employed using secondary data from 33 districts and cities in North Sumatra for the year 2023. Factor analysis, including Principal Component Analysis (PCA), was used to reduce variable dimensionality, while K-Means Clustering grouped regions with similar socio-economic profiles. Regression analysis was then applied to test the statistical significance of key factors influencing poverty. Five principal components were identified as significant poverty determinants: (1) education and human capital, (2) access to basic sanitation, (3) infrastructure and regional economic development, (4) employment rate, and (5) local fiscal capacity. The clustering procedure revealed five distinct regional typologies, ranging from urbanized, economically advanced areas to underdeveloped, rural localities. Education and unemployment emerged as the most consistent predictors of poverty levels across clusters. The findings provide a strong empirical foundation for spatially differentiated poverty alleviation strategies. Policymakers are encouraged to abandon the "one-size-fits-all" approach and adopt geographically nuanced interventions. This study advances the discourse on spatial poverty by combining multivariate statistical techniques with regional clustering to reveal intra-provincial heterogeneity. It provides actionable insights for regional development planning in decentralized governance contexts.

**Keywords**: Poverty, Spatial Clustering, Factor Analysis, North Sumatra, Regional Inequality, Multidimensional Poverty, K-Means, PCA

1. **INTRODUCTION**

Poverty remains one of the most pressing global challenges, particularly in developing countries like Indonesia, where it intertwines with issues of inequality, education, health, and regional development. Despite measurable progress in reducing national poverty rates, approximately 9.78% of Indonesians still lived below the poverty line as of 2020 [(Purwanti, 2023)](https://scispace.com/papers/clustering-analysis-of-multidimensional-poverty-in-central-1bdkhtaw7u?utm_source=chatgpt). The multidimensional nature of poverty, including economic, social, and infrastructural deficiencies, necessitates more granular and region-specific policy interventions. Empirical studies have shown that education, unemployment, and public health budgets significantly influence poverty dynamics, often exhibiting regional variation in their impact [(Sulaeman et al., 2024)](https://scispace.com/papers/economic-growth-in-indonesia-dynamics-of-poverty-37raj94yf1la?utm_source=chatgpt).

Indonesia, as an archipelagic nation, is marked by stark regional disparities. Provinces like North Sumatra exhibit wide intraregional variations in poverty due to unequal access to basic services, geographic constraints, and labor market differences. For instance, some regencies face chronic underdevelopment, while others benefit from industrialization or better infrastructure (Wiguna et al., 2023). This spatial heterogeneity makes poverty not only a national concern but also a critical regional development issue. Moreover, causality analyses in East Java have shown that education and unemployment maintain significant bidirectional relationships with poverty and inequality, further reinforcing the need for localized studies [(Firdauzi & Dewi, 2022)](https://scispace.com/papers/analysis-of-causality-interactions-between-education-3m8jyt79?utm_source=chatgpt).

Conventional “one-size-fits-all” poverty reduction strategies have often proven ineffective, especially in a country with such geographic and socioeconomic diversity. The adoption of uniform interventions can overlook local poverty drivers and lead to inefficient resource allocation. Recent spatial econometric analyses highlight the need for localized strategies tailored to specific regional characteristics (Khamila et al., 2024). For example, studies comparing clustering algorithms like K-Medoids and CLARA reveal that algorithmic choice can significantly affect poverty mapping, especially in areas with overlapping social indicators [(Ardini & Sirait, 2023)](https://scispace.com/papers/comparison-of-k-medoids-and-clara-algorithm-in-poverty-39jaifudbz?utm_source=chatgpt).

Moreover, recent spatial panel data analyses have emphasized that the availability of education and health infrastructure—such as school buildings, health centers, and sanitation—plays a significant role in regional poverty levels. Villages farther from urban centers or lacking basic facilities report higher poverty incidence, especially in rural and topographically challenging areas (Rahayu et al., 2020). Similarly, a study on infrastructure accessibility across Indonesia found that while educational and healthcare investments reduce poverty significantly, road infrastructure showed weaker impacts unless integrated with service delivery and employment access (Arma et al., 2018).

The spatial dependence of poverty has also been documented in coastal and semi-urban regions like the Pantura area of East Java, where spatial regression models revealed interconnected effects of natural resource access, education, and regional connectivity on poverty clustering (Ifa et al., 2024). These insights support a growing consensus: effective poverty interventions must consider both spatial autocorrelation and regional structural contexts. Furthermore, the availability of infrastructure like electricity, sanitation, and high schools significantly correlates with lower poverty in spatial models across Indonesia [(Pramono & Marsisno, 2018)](https://scispace.com/papers/availability-of-infrastructure-for-poverty-reduction-in-4iogmy6oic?utm_source=chatgpt).

This research addresses these limitations by applying a clustering-based analysis to classify districts in North Sumatra based on shared poverty determinants. Through statistical modeling and pattern recognition, the study aims to support more context-sensitive planning and policy targeting. The literature on poverty determinants frequently emphasizes variables such as economic growth, education, housing conditions, and unemployment. Studies have found that open unemployment, inadequate housing, and poor access to sanitation strongly correlate with poverty severity in Indonesian provinces [(Purwanti, 2023); (Wiguna et al., 2023); (Anggraeni, 2022)](https://scispace.com/papers/pengentasan-angka-kemiskinan-di-indonesia-tahun-2015-2019-2ap6yx27?utm_source=chatgpt).

A growing body of literature uses clustering algorithms like K-Means and Hierarchical Clustering to uncover regional groupings based on socioeconomic indicators. These methods have proven effective in classifying districts into poverty risk categories, thus guiding targeted interventions (Annas et al., 2022). Cluster analysis also facilitates cross-regional comparisons and the identification of spatial outliers (Khamila et al., 2024). However, existing studies tend to concentrate on national or Java-based contexts, with fewer empirical explorations focusing on Sumatra. Moreover, most clustering analyses rely solely on income or consumption-based metrics, neglecting multidimensional aspects such as education or infrastructure quality (Setiawan & Zahra, 2023).

This thesis contributes by filling this gap—integrating diverse poverty determinants into a spatial cluster analysis specifically for North Sumatra. By incorporating socio-economic, infrastructural, and policy variables into the clustering model, this study builds on previous spatial regression work while offering actionable insights for local governments. It enhances the understanding of poverty as a spatial phenomenon and demonstrates how data-driven grouping can improve poverty targeting at the sub-provincial level.

**II. METHODS**

**2.1 Research Design**

This study employs a quantitative approach with descriptive and cluster analysis techniques to identify spatial patterns and determinants of poverty across regencies and cities in North Sumatra. The analysis is designed to provide an empirical basis for localized poverty reduction policies through a data-driven regional classification.

**2.2 Data and Variables**

The data used were secondary and drawn primarily from Badan Pusat Statistik (BPS) publications for the year 2023. The unit of analysis includes 33 districts/cities in North Sumatra. The dependent variable is the poverty rate (percentage of poor population), while independent variables include:

* Economic growth (GRDP growth rate)
* Open unemployment rate
* Minimum wage (UMK)
* Local government revenue (PAD)
* Education level (average years of schooling)
* Housing conditions (percentage of uninhabitable housing)

All variables were standardized to Z-scores prior to further analysis to ensure comparability across different scales of measurement.

**2.3 Analytical Methods**

a) Descriptive Analysis

Descriptive statistics were used to analyze central tendency (mean), dispersion (standard deviation), and distribution characteristics (skewness, kurtosis) for each variable across 33 districts/cities. This stage also supported outlier detection and normalization needs prior to multivariate analysis.

b) Factor Analysis (Principal Component Analysis - PCA)

PCA was utilized to reduce the dimensionality of the dataset by extracting latent factors from intercorrelated variables. The PCA transformation is defined as:

Z=XA

Where:

* Z = matrix of principal components
* X = standardized data matrix
* A = eigenvector matrix of the covariance matrix of XXX

The Kaiser-Meyer-Olkin (KMO) measure and Bartlett’s Test of Sphericity were used to assess data adequacy. Factors with eigenvalues > 1 and cumulative variance ≥ 60% were retained for further analysis.

c) Cluster Analysis (K-Means)

K-Means clustering aims to partition nnn observations into kkk clusters C1,C2,...,CkC\_1, C\_2, ..., C\_kC1​,C2​,...,Ck​ by minimizing within-cluster variance. The objective function is:



Where:

* xj ​ = data point
* μi​ = centroid of cluster i
* ∥xj−μi∥2= Euclidean distance between xj​ and cluster center μi

The optimal number of clusters k was determined using:

* Elbow Method: identifies the point at which the marginal gain in intra-cluster variance reduction drops
* Silhouette Coefficient: measures how similar a point is to its own cluster compared to other clusters

Where:

* a(i): average distance of iii to all points in its cluster
* b(i): minimum average distance of iii to points in other clusters

d) Regression Analysis

To quantify the effect of independent variables on poverty rate, a multiple linear regression model was used:

Y=β0+β1X1+β2X2+...+βkXk+ϵ

Where:

* Y = dependent variable (poverty rate)
* β0 ​ = intercept
* βk ​ = regression coefficients
* Xk ​ = independent variables (e.g., education, unemployment)
* ϵ = error term

Assumptions were checked using:

* Multicollinearity: Variance Inflation Factor (VIF)
* Heteroscedasticity: Breusch-Pagan test
* Normality of residuals: Kolmogorov–Smirnov or Shapiro–Wilk test

**III. RESULTS AND DISCUSSION**

**3.1 Descriptive Statistical Insights**

The descriptive statistics indicated pronounced variability in poverty-related indicators across districts. Education, for instance, showed that districts like Medan and Binjai had averages exceeding 9.5 years of schooling, while some outer islands, such as Nias Barat, had averages below 6.5 years. This reflects educational inequality, often driven by geographic isolation and infrastructure deficits.

Similarly, **unemployment rates** ranged widely. Urban centers exhibited moderate unemployment (5–7%), whereas rural and remote areas showed both extremes—some with minimal formal sector engagement and others with chronic joblessness. The high variance underscores the need to understand regional labor dynamics rather than applying uniform employment policies.

**Minimum wages (UMK)** varied not only by geographic location but also by regional economic bases. Deli Serdang had among the highest minimum wages due to industrial concentration, while island districts with limited economic activities posted lower figures. These discrepancies mirror the uneven distribution of formal employment opportunities.

Housing conditions further accentuated regional disparities. Districts such as Padang Lawas Utara had up to 27.6% of houses classified as uninhabitable, while urban areas showed better conditions. Housing is often linked with public investment and local governance effectiveness.

**3.2 Factor Extraction via PCA**

PCA reduced the complexity of the dataset by condensing multiple variables into three principal components:

* **Component 1 (Development Capital)**: included PAD, education level, and economic growth. It reflects the systemic development capacity of each district.
* **Component 2 (Labor Market Risk)**: included unemployment and wage indicators. It encapsulates workforce vulnerability.
* **Component 3 (Housing Deprivation)**: focused on poor housing infrastructure, interpreted as a proxy for multidimensional poverty.

These three components captured over 68% of total variance, indicating a strong explanatory power and confirming that poverty in North Sumatra is both economically and socially multidimensional.

**3.3 Cluster Analysis and Regional Typology**

Using K-Means clustering, five distinct clusters were derived. The Silhouette Coefficient (~0.62) confirmed moderately strong cohesion and separation of clusters, validating the chosen segmentation.

**Figure 1.** Spatial Distribution of Poverty Levels in North Sumatra (2023)

**Legend (Poverty Level):**

* **Very High** (Dark Brown)
* **High** (Brown)
* **Moderate** (Orange)
* **Low** (Light Orange)
* **Very Low** (Yellow)

**Description**:
The map displays the spatial distribution of poverty intensity across 33 districts and cities in North Sumatra. Island regencies like Nias (NIU, NIA, NSL) exhibit very high poverty, while urban centers such as Medan (MDN) and Deli Serdang (DLS) show very low to low levels of poverty. This visual pattern reinforces the findings from the cluster analysis, highlighting stark disparities tied to infrastructure, regional income, and public service access.

Table 1. Cluster Analysis

|  |  |  |
| --- | --- | --- |
| **Cluster** | **Characteristics** | **District Examples** |
| C1 | Low poverty, high PAD, good education | Medan, Binjai |
| C2 | High poverty, low infrastructure | Nias, Nias Barat |
| C3 | Medium poverty, strong education but weak labor market | Tebing Tinggi |
| C4 | Balanced indicators with moderate vulnerabilities | Langkat |
| C5 | High unemployment, low wage | Mandailing Natal |

* **Cluster 1 (Urban Advantage)**: Comprised largely of metropolitan areas (e.g., Medan, Binjai) with high education, low unemployment, strong PAD, and adequate housing. These areas have effectively leveraged their infrastructure and economic base to suppress poverty levels.
* **Cluster 2 (Rural Deprivation)**: Included regencies like Nias and Nias Barat. These areas are marked by extremely low PAD, high housing deprivation, and underdeveloped human capital. Poverty here is structural and rooted in geographic isolation and limited fiscal capacity.
* **Cluster 3 (Mismatch Syndrome)**: Represented districts with decent educational access but poor employment outcomes, such as Tebing Tinggi. This mismatch between education and job creation points to potential problems in labor absorption and industrial connectivity.
* **Cluster 4 (Moderate Development)**: Showed balanced performance on all indicators but lacked specialization. Districts like Langkat fall here, often performing “average” on all variables without any standout weaknesses or strengths.
* **Cluster 5 (Wage Trap)**: Consisted of districts with low PAD, high unemployment, and underwhelming wage structures, such as Mandailing Natal. These areas may have some industrial activity but remain plagued by labor underutilization and poor job quality.

**3.4 Interpretative Implications**

This typology is not merely academic—it suggests tailored interventions:

* For Cluster 1, continued investment in innovation, higher education, and urban infrastructure will sustain low poverty levels.
* For Cluster 2, priority should be basic infrastructure and connectivity—particularly in health, education, and logistics.
* Cluster 3 needs labor market alignment, such as vocational training, industry incentives, and job-matching mechanisms.
* In Cluster 4, a strategy of diversification and upgrading public services may unlock latent growth potential.
* Cluster 5 requires minimum wage enforcement, SME development, and social safety net enhancements.

**3.5 Regression Model Findings**

The regression model confirmed several critical factors:

* **Education level** was the strongest negative predictor of poverty (β=−0.33,p<0.01\beta = -0.33, p < 0.01β=−0.33,p<0.01). This supports long-standing literature emphasizing education's role in upward mobility and labor productivity.
* **Unemployment** (β=+0.29\beta = +0.29β=+0.29) emerged as a significant poverty driver, reflecting that lack of income opportunity leads directly to deprivation.
* **PAD** (β=−0.26\beta = -0.26β=−0.26) reinforces the importance of regional fiscal capacity—districts with higher own-source revenue can better support welfare programs.
* **Uninhabitable housing** (β=+0.22\beta = +0.22β=+0.22) was also statistically significant, aligning with recent findings that link poor housing to chronic health problems and social exclusion.

Interestingly, **minimum wage** was not a significant predictor in the final model, likely due to high informal employment in several districts. This supports research showing that wage regulations in Indonesia often do not extend to the informal sector (Setiawan & Zahra, 2023)

**3.6 Comparative Reflection with Previous Studies**

Compared to regional poverty studies conducted in Java, such as the work by Purwanti (2023), which utilized clustering methods on Central Java, this study contributes a more geographically diverse and context-sensitive perspective by focusing on North Sumatra, a province with distinct topographical, cultural, and fiscal features. Java-based studies, while methodologically rigorous, often focus on areas with relatively better infrastructure, institutional capacity, and access to services, which limits their generalizability to more diverse or remote regions. North Sumatra, by contrast, encompasses a wide spectrum of development conditions—from metropolitan cities like Medan to isolated islands such as Nias—making it an ideal case to showcase intra-provincial disparities in poverty dynamics. This finer granularity allows the present research to unearth region-specific vulnerabilities that might be masked in national or Java-centric analyses.

Furthermore, in contrast to studies like Wiguna et al. (2023), which apply spatial regression to explain national-level poverty trends using macro indicators, this study enhances explanatory power by integrating Principal Component Analysis (PCA), K-Means clustering, and regression modeling into a unified framework. While spatial regression is effective for identifying variable associations across regions, it may overlook latent patterns and natural groupings within the data. By applying PCA to reduce multicollinearity and dimensionality, followed by clustering to classify regencies based on shared characteristics, this study not only identifies poverty correlates but also delineates actionable cluster typologies. This methodological triangulation represents a significant advancement over traditional linear-only or unsupervised-only approaches, especially in complex, heterogeneous socio-economic landscapes like North Sumatra.

Perhaps most notably, this study contributes conceptually by moving beyond single-dimensional, income-based poverty metrics commonly used in national surveys. While income remains a crucial measure, it is increasingly recognized—by institutions such as UNDP and Indonesia’s Central Statistics Agency (BPS)—that poverty is a multidimensional phenomenon encompassing health, education, living standards, and infrastructure. By including variables such as housing conditions, education attainment, and local fiscal capacity (PAD), this research aligns more closely with the Multidimensional Poverty Index (MPI) framework. The inclusion of uninhabitable housing as a proxy for living condition deprivation, for example, captures material hardship not reflected in income data alone. This shift ensures a more holistic understanding of deprivation, offering richer insights for policy formulation and program targeting at the local level.

**3.7 Policy Relevance and Limitations**

This research supports **localized targeting** under Indonesia's decentralization framework. Regional development planning (RPJMD) can incorporate such clustering to allocate budgets more effectively.

However, the study has limitations: it uses cross-sectional data from 2023, so temporal dynamics are not captured. Also, informal sector employment and gendered poverty dimensions were not directly addressed.

**IV. CONCLUSION AND POLICY RECOMMENDATIONS**

This study offers a comprehensive and spatially nuanced examination of poverty across North Sumatra by employing a blend of descriptive statistics, Principal Component Analysis (PCA), K-Means clustering, and multiple regression techniques. By adopting a multidimensional perspective—including education, unemployment, regional income (PAD), housing quality, and minimum wage—the study highlights complex patterns and drivers of poverty at the district level. The analysis resulted in the identification of five distinct poverty clusters, ranging from urban centers with robust socio-economic conditions to rural and remote regions facing entrenched deprivation due to limited infrastructure and weak fiscal capacity.

Each cluster reflects a unique typology of poverty, revealing the inadequacy of generalized, one-size-fits-all policies. For instance, urban districts such as Medan and Binjai fall into a "low poverty–high capacity" group that benefits from strong institutional and educational ecosystems, while regions like Nias Barat experience compounded structural challenges including poor housing and low PAD. Importantly, the regression analysis confirmed that education, unemployment, housing conditions, and regional fiscal resources are statistically significant predictors of poverty. The effect of minimum wage was less impactful, likely due to the dominance of informal employment in many districts. These findings provide clear evidence for the necessity of locally tailored policy strategies that consider regional disparities in socioeconomic conditions.

Accordingly, the study recommends differentiated interventions aligned with each cluster’s profile. Urban districts should focus on maintaining their growth momentum through digital infrastructure and inclusive urban policies. Remote areas require basic infrastructure investment and better access to public services. Districts facing skill-job mismatches need targeted vocational training and stronger industry partnerships. Additionally, cluster-based approaches should be formally integrated into regional planning (RPJMD), complemented by inter-district collaboration frameworks and improved data monitoring systems, particularly for informal sector dynamics. This model not only enhances policy precision in North Sumatra but also offers a scalable approach for other provinces facing similar heterogeneity in poverty profiles.

**References**

Anggraeni, R. (2022). Poverty alleviation in Indonesia in 2015–2019. Transekonomika: Accounting, Business and Finance, 2(4). https://doi.org/10.55047/transekonomika.v2i4.136

Annas, M., Komara, R., & Novanti, L. R. (2022). Regional Clustering Based on Poverty Indicators Using the Hierarchical Agglomerative Clustering Algorithm. Matrik: Journal of Management, Business Strategy and Entrepreneurship, 21(2), 183–192. https://doi.org/10.30812/matrik.v21i2.1289

Ardini, A. R. D., & Sirait, H. (2023). Comparison of K-Medoids and CLARA Algorithm in Poverty Clustering Analysis in Indonesia. ORICS, 4(4). https://scispace.com/papers/comparison-of-k-medoids-and-clara-algorithm-in-poverty-39jaifudbz?utm\_source=chatgpt

Firdauzi, I., & Dewi, N. M. R. (2022). Analysis of causality interactions between education, inequality, and unemployment toward poverty in East Java: Empirical evidence from dynamic panel co-integration model. Journal of Economics Research and Social Sciences, 6(1). https://doi.org/10.18196/jerss.v6i1.13568

Ifa, K., Viphindrartin, S., Santoso, E., & Priyono, T. H. (2024). Spatial dependence and poverty factors: A study of Pantura Regions in East Java Province, Indonesia. Journal of Infrastructure, Policy and Development, 8(7). https://doi.org/10.24294/jipd.v8i7.4195

Khamila, A., Ibrahim, M., & Mulya, S. (2024). Poverty Data Clustering Using DBSCAN and K-Means Algorithms Based on Silhouette Index Performance. Journal of Informatics and Data Analysis Mandiri, 7(1), 13–21. https://doi.org/10.24014/ijaidm.v7i1.25278

Pramono, G., & Marsisno, W. (2018). Availability of Infrastructure for Poverty Reduction in Indonesia: Spatial Panel Data Analysis. Economics and Finance in Indonesia, 64(2), 145–166. https://doi.org/10.7454/EFI.V64I2.587

Purwanti, Y. (2023). Clustering analysis of multidimensional poverty in Central Java Province, Indonesia. Jurnal Inovasi Daerah, 2(2). https://doi.org/10.56655/jid.v2i2.132

Rahayu, H. C., Sarungu, J. J., Hakim, L., Soesilo, A. M., Lestari, E. P., Astuti, D., & Retnaningsih, T. K. (2020). Dynamic Panel Data Analysis of Poverty in Indonesia. Proceedings of the 1st Economics and Business International Conference 2020. https://doi.org/10.2991/AEBMR.K.200522.028

Setiawan, B. D., & Zahra, F. F. (2023). Modeling Provincial Poverty in Indonesia Using Spatial Autoregressive Model (SAR). Jurnal Statistika dan Sains Data, 6(1), 12–24. https://doi.org/10.12962/j27213862.v6i1.14969

Sulaeman, M., Suharno, S., & Ahmad, A. A. (2024). Economic growth in Indonesia: Dynamics of poverty, unemployment, and health budgets. Agregat: Jurnal Ekonomi dan Bisnis, 8(2). https://doi.org/10.22236/agregat\_vol8.i2/16418

Wiguna, A. C., Savitri, Y., & Ketut, I. A. C. (2023). Spatial regression and spatial autocorrelation analysis of the determinants of poverty in Indonesia in 2022. Jurnal Info Artha, 7(2), 21–30. https://doi.org/10.31092/jia.v7i2.2318