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

**Abstract :**

This study investigates the multidimensional nature of poverty and its spatial distribution across 33 districts and cities in North Sumatra, Indonesia. Using a combination of Principal Component Analysis (PCA) and K-Means clustering, the research identifies key socioeconomic factors contributing to regional poverty disparities, including education, unemployment, housing quality, and local fiscal capacity. The clustering results reveal five distinct district typologies, ranging from urban centers with strong infrastructure and human capital to rural and island regions facing structural deprivation. Multiple regression analysis confirms that education level, unemployment, and uninhabitable housing significantly predict poverty levels. The findings highlight the limitations of one-size-fits-all poverty alleviation strategies and underscore the need for geographically targeted policies. This study provides empirical evidence to support region-specific planning under Indonesia’s decentralized governance framework and offers a scalable approach for other provinces facing similar socio-economic diversity.

**Keywords**: multidimensional poverty, spatial clustering, regional inequality, factor analysis, K-means, North Sumatra, poverty policy

1. **INTRODUCTION**

Poverty continues to be a multidimensional challenge that affects millions across Indonesia, particularly in regions where development disparities remain acute. In North Sumatra, persistent poverty rates are intricately tied to structural factors such as education, infrastructure, regional governance, and labor market conditions (Gea, 2023; Sitanggang et al., 2024). Despite various national strategies and social interventions, including the Strategic Poverty Alleviation Plan 2021–2025 (Bappenas, 2021), poverty in the province remains unevenly distributed, reflecting spatial inequalities and localized development gaps.

Multiple studies have underscored that regional poverty is not merely an economic issue but a composite of social exclusion, limited human capital, and infrastructural deficiencies (Afifi & Nasir, 2023; Amin et al., 2024). For instance, empirical evidence reveals that the allocation of village funds and local social assistance has accelerated poverty alleviation in some areas but failed to reach structural depth in others (Rambe et al., 2023). The quality of human capital, proxied through education levels and employment readiness, also shows strong associations with regional poverty rates (Abrar et al., 2024; Nasution et al., 2024).

In North Sumatra, regency-level disparities highlight the critical need for disaggregated poverty diagnostics. A recent analysis found that indicators such as population growth, access to education, and employment opportunities significantly shape poverty intensity in both rural and semi-urban districts (Sinambela et al., 2024; Hidayah, 2024). Moreover, studies using spatial econometric models argue that the proximity of economic infrastructure—roads, schools, and health facilities—directly influences the poverty distribution patterns across the province (Aldawiyah et al., 2024; Fitriawan et al., 2023).

Cluster analysis has emerged as a vital method in understanding such multidimensional phenomena. Techniques like K-Means and Principal Component Regression (PCR) have been used to classify poverty-prone regions based on composite indicators, enabling more nuanced policy responses (Suryani et al., 2023; Yuningsih & Aryani, 2024). In particular, recent statistical applications have emphasized that clustering districts based on shared social and economic characteristics can uncover spatial poverty traps and highlight development blind spots (Amelia et al., 2023; Ardiansyah & Ilyas, 2023).

However, poverty mapping efforts often rely solely on income thresholds, which overlook non-monetary dimensions such as housing quality, education access, and health conditions (Ilahi et al., 2024). This leads to policy gaps, especially in provinces like North Sumatra, where topographical and cultural diversity require more context-aware interventions. Studies in other Indonesian provinces have successfully demonstrated that integrating education, unemployment, and health indicators into spatial clustering leads to more effective poverty classification and program targeting (Adam et al., 2022; Aditya & Wahed, 2021).

Furthermore, the interaction between governance performance and human capital development has been shown to influence the sustainability of poverty reduction efforts. Poor local governance, limited institutional capacity, and weak inter-agency coordination continue to obstruct the full realization of anti-poverty initiatives (Ali et al., 2022; Arifah & Sunarjo, 2021). These findings call for a multidimensional and regionally adaptive approach to poverty research.

Building on this literature, the present study applies a clustering-based approach to classify the districts in North Sumatra according to multidimensional poverty indicators. By incorporating socio-economic, infrastructural, and governance-related variables, this research offers a more granular understanding of poverty dynamics at the sub-provincial level. It aims to support policymakers in designing tailored interventions that are sensitive to regional variations and rooted in data-driven insights.

**II. METHODS**

This study employs a mixed quantitative approach to analyze multidimensional poverty across 33 districts and cities in North Sumatra Province, Indonesia. The methodology integrates **Principal Component Analysis (PCA)**, **K-Means clustering**, and **multiple linear regression** to identify the key determinants of poverty and spatial typologies that can inform region-specific interventions.

**2.1 Data Sources and Variables**

The study uses **secondary data** from the **Badan Pusat Statistik (BPS)** for the year **2023**, encompassing variables that reflect the economic, educational, labor, housing, and fiscal dimensions of poverty. The dependent variable is the **poverty rate (%)** per district. The independent variables include:

* **Education Attainment Rate (% of ≥Senior High School graduates)**
* **Open Unemployment Rate (%)**
* **Percentage of Uninhabitable Housing (%)**
* **Per Capita Regional Revenue (PAD, in IDR)**
* **Minimum Regional Wage (UMR, in IDR)**

Prior to analysis, all variables were **standardized using Z-scores** to remove scale effects and ensure comparability across indicators. This normalization is essential for PCA, as it prevents variables with larger magnitudes from dominating the factor extraction process.

**2.2 Principal Component Analysis (PCA)**

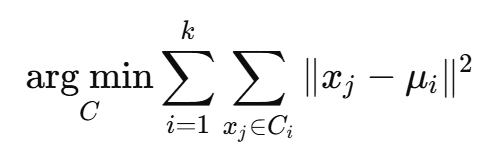
PCA was employed to reduce dimensionality and extract latent structures among the correlated socioeconomic indicators. The **Kaiser-Meyer-Olkin (KMO)** test yielded a value of **0.694**, indicating sampling adequacy, while **Bartlett’s Test of Sphericity** was significant (**p < 0.01**), validating the factorability of the dataset. Based on **eigenvalues > 1.0** and the **scree plot**, **three principal components** were extracted:

* **Component 1**: Development Capital (Education, PAD, UMR)
* **Component 2**: Labor Market Vulnerability (Unemployment)
* **Component 3**: Housing Deprivation (Uninhabitable Housing)

These components were used as inputs for clustering to reveal multidimensional patterns across districts.

**2.3 K-Means Clustering**

To classify districts into poverty typologies, **K-Means clustering** was conducted using the standardized scores of the three principal components. The optimal number of clusters (**k = 5**) was determined using a combination of the **Elbow method** and **Silhouette analysis**. Each resulting cluster represents a group of districts with shared structural poverty characteristics:



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

The clustering analysis was visualized spatially using **QGIS**, with districts color-coded by cluster membership to highlight geographical disparities and groupings.

**2.4 Multiple Linear Regression**

A regression model was estimated to quantify the effect of each original socioeconomic variable on the poverty rate. The **ordinary least squares (OLS)** method was used.

Y=β0+β1X1+β2X2+...+βkXk+ϵ

Diagnostic tests were performed:

* **Variance Inflation Factor (VIF)** confirmed no multicollinearity (all < 5).
* **Breusch-Pagan test** showed no heteroscedasticity (**p > 0.05**).
* **Shapiro-Wilk test** supported the normality of residuals.

This step validated the explanatory power of the variables selected in PCA and cluster analysis and allowed the identification of statistically significant poverty determinants in the region

Where:

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

**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.

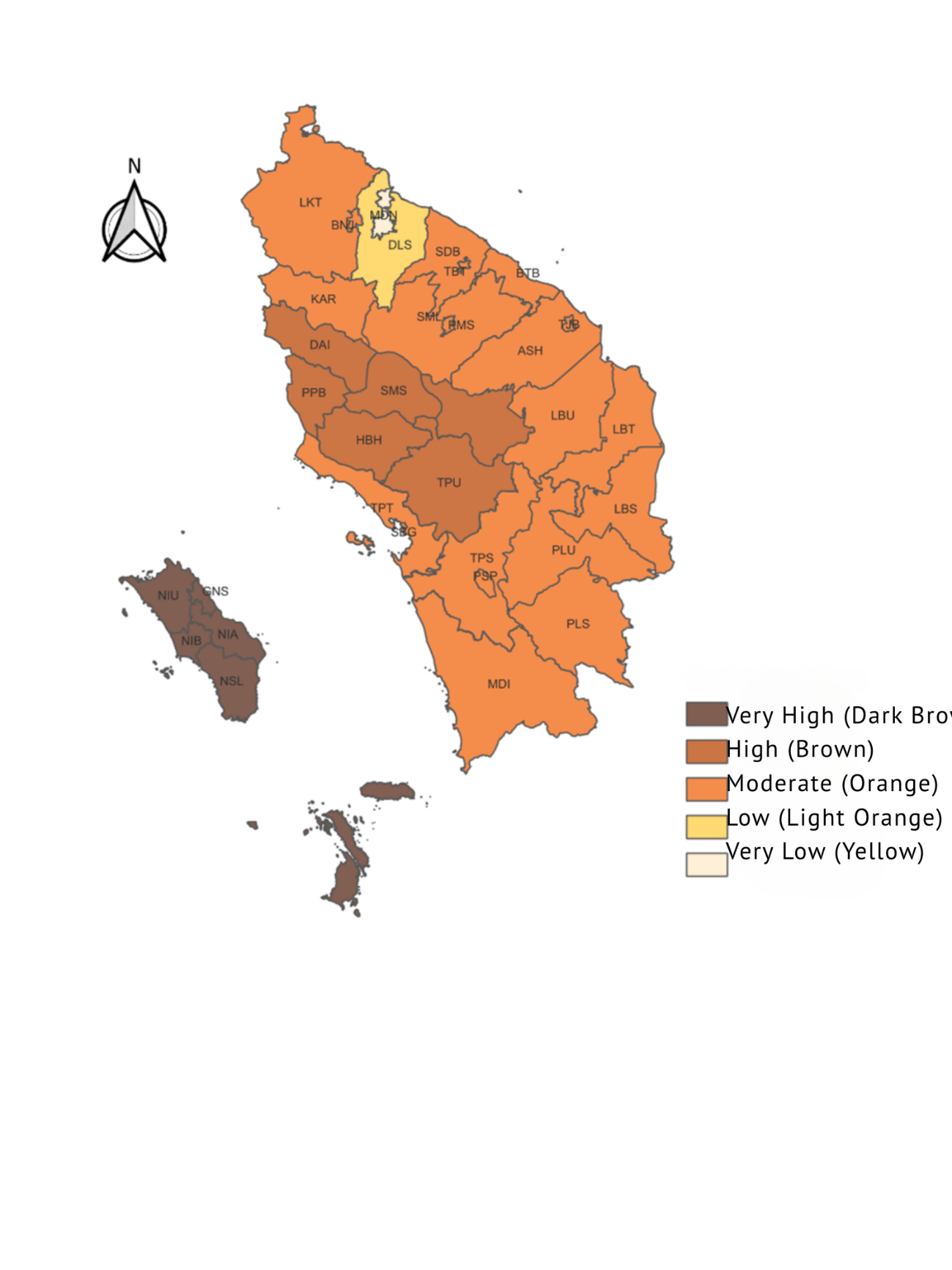
**3.2 Factor Extraction via PCA**

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**



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

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.

**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.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

**References**

**Abrar, M., Sriani, S., & Harahap, N. (2024). *Information system for new student admissions in the SMA development sector, Department of Education, North Sumatra Province*. Journal of Computer and Business Science, 15(2), 174–183.** [**https://doi.org/10.47927/jikb.v15i2.810**](https://doi.org/10.47927/jikb.v15i2.810)

**Adam, D., Olilingo, F., & Santoso, I. (2022). *Analysis of the effect of education and unemployment on poverty in the Northern-Northern Cooperation Area*. Journal of Development Economics, 8(1).** [**https://doi.org/10.35906/jep.v8i1.1039**](https://doi.org/10.35906/jep.v8i1.1039)

**Aditya, I., & Wahed, M. (2021). *The effect of local revenue and unemployment on the human development index with poverty as an intervening variable in Bali Province*. Randai Journal, 2(1), 1–10.** [**https://doi.org/10.31258/randai.2.1.p.1-10**](https://doi.org/10.31258/randai.2.1.p.1-10)

**Afifi, M., & Nasir, M. (2023). *Distribution of village funds and the acceleration of poverty alleviation in Lombok Island*. Aliansi Journal of Sharia Accounting and Finance, 6(2), 124–131.** [**https://doi.org/10.54712/aliansi.v6i2.284**](https://doi.org/10.54712/aliansi.v6i2.284)

**Ali, M., Egbetokun, A., & Memon, M. (2022). *Human capital and economic growth: The role of governance*. The Pakistan Development Review, 61(4), 529–549.** [**https://doi.org/10.30541/v54i4i-iipp.529-549**](https://doi.org/10.30541/v54i4i-iipp.529-549)

**Aldawiyah, N., Astuti, A., Kurnia, R., Amalia, N., Amelia, D., Mardianto, M., … & Ana, E. (2024). *Analysis of poverty factors in North Sumatra Province using principal component regression*. Variance Journal of Statistics and Its Applications, 6(1), 63–74.** [**https://doi.org/10.30598/variancevol6iss1page63-74**](https://doi.org/10.30598/variancevol6iss1page63-74)

**Amelia, A., Nur, I., Rizky, M., & Milasari, S. (2023). *Poverty level grouping in West Java Province using the K-Means clustering method*. Journal of Data Insights, 1(2), 51–61.** [**https://doi.org/10.26714/jodi.v1i2.152**](https://doi.org/10.26714/jodi.v1i2.152)

**Amin, C., Zamzam, I., Reviane, I., Duko, F., Hasnin, M., Muhammad, N., … & Anwar, C. (2024). *Performance of social assistance on local economic development: Evidence from island regions in Eastern Indonesia*. Journal of Economics and Development Studies, 25(1), 97–109.** [**https://doi.org/10.18196/jesp.v25i1.20988**](https://doi.org/10.18196/jesp.v25i1.20988)

**Ardiansyah, A., & Ilyas, A. (2023). *Harmonization of Indonesia’s integrated poverty data policy aligned with the Sustainable Development Goals*. Journal of Law and Sustainable Development, 11(11), e1413.** [**https://doi.org/10.55908/sdgs.v11i11.1413**](https://doi.org/10.55908/sdgs.v11i11.1413)

**Arifah, L., & Sunarjo, D. (2021). *Analysis of inter-industry linkages in North Sumatra and their impact on the Indonesian economy in 2016 (IO and IRIO analysis)*. Journal of Economics and Statistics Indonesia, 1(3), 213–226. https://doi.org/10.11594/jesi.01.03.08**

**Fitriawan, M., Risada, N., Yaseva, Y., & Putri, R. (2023). *Analysis of agricultural land carrying capacity in North Sumatra Province in 2020*. Bio Web of Conferences, 80, 02001.** [**https://doi.org/10.1051/bioconf/20238002001**](https://doi.org/10.1051/bioconf/20238002001)

**Gea, K. (2023). *Factors affecting poverty in North Sumatra Province*. Journal of Economics and Strategy, 4(1), 115–123.** [**https://doi.org/10.36490/jes.v4i1.747**](https://doi.org/10.36490/jes.v4i1.747)

**Hidayah, N. (2024). *Impact analysis of education, housing, and unemployment on poverty levels in districts/cities in East Java*. Journal of Economic Management and Accounting Science (MEA), 8(1), 2095–2104.** [**https://doi.org/10.31955/mea.v8i1.3907**](https://doi.org/10.31955/mea.v8i1.3907)

**Ilahi, S., Restuhadi, F., & Rifai, A. (2024). *Analysis of factors affecting poverty on Sumatra Island using bootstrapping approach in SEM-PLS*. Indonesian Journal of Agricultural Economics, 14(1), 32–49.** [**https://doi.org/10.31258/ijae.14.1.32-49**](https://doi.org/10.31258/ijae.14.1.32-49)

**Sinambela, H., Suryani, I., Sitinjak, R., Pasaribu, M., & Sihombing, T. (2024). *Effect of population, human development index, and unemployment on poverty in North Sumatra from 2003 to 2022*. Journal of Management and Administration, 2(2).** [**https://doi.org/10.62281/v2i2.174**](https://doi.org/10.62281/v2i2.174)