SINOMICS JOURNAL
International Journal of Social Science, Education, Communication and Economics
ISSN (e): 2829-7350 | ISSN(p): 2963-9441

Implementation of K-Means Clustering in Poverty Analysis of Regency/City in Sumatera Island in 2023

Siti Nur Fadilah[1]*, Dedy Yuliawan[2]
University of Lampung, Indonesia
***E-mail Correspondence:** sitinurfadilah152@gmail.com

Abstract

This study seeks to examine poverty across the regencies and cities on Sumatra Island in 2023 by employing the K-Means Clustering approach. Poverty represents a complicated and multi-faceted societal challenge, shaped by various elements including educational attainment, joblessness, income per capita, and spending per capita. The information utilized in this analysis is sourced from the Central Bureau of Statistics, specifically the proportions of impoverished individuals, average duration of education, rates of open unemployment, income per capita, and expenditure per capita. Findings reveal the establishment of three distinct clusters based on poverty attributes: Cluster 1 exhibiting a low poverty level, Cluster 2 displaying a moderate poverty level, and Cluster 3 indicating a high poverty level. Results from the One Way Anova test indicate notable differences in poverty traits across the clusters. It is anticipated that this research will aid the government in developing more suitable and effective strategies for tackling poverty in regions grappling with significant poverty challenges.

Keywords *Poverty, K-Means Clustering, One Way Anova.*

INTRODUCTION

Poverty presents a complicated and multifaceted societal concern, signifying a condition where individuals or groups lack the necessary means or chances to meet life's essential requirements. In Indonesia, poverty stands out as a prominent hurdle for the populace, particularly in regions identified as developing. Despite the government's numerous attempts to tackle this issue, the outcomes have not been entirely satisfactory, and Indonesia is still viewed as a developing nation. It is imperative to initiate efforts aimed at addressing the poverty dilemma, making poverty alleviation one of the vital components for attaining the well-being of many citizens at both national and local levels (Mayasari & Nugraha, 2023). Sumatra Island, noted as one of Indonesia's most significant islands, encounters its own set of obstacles concerning poverty. The World Bank indicates that Sumatra Island is identified as an area ready for development as a growth hub capable of attracting investments and resources for economic growth, with the potential to outperform Java Island in economic advancement and development (Salsabila, 2020). Spanning roughly 443,065.8 km², Sumatra Island is among Indonesia's largest islands, its advantageous location and abundant natural resources fostering various economic endeavors (Amalia & Emalia, 2022). Despite the island's substantial economic capabilities, numerous provinces and regions in Sumatra continue to struggle with notable poverty challenges.

The city with the smallest proportion of impoverished individuals in Sumatra is Sawahlunto City, with just 2.27 percent, followed closely by Bangka Barat at 2.71 percent. These regions, having the minimal figures of poverty, could serve as examples for other provinces in devising and executing poverty reduction strategies and initiatives by examining the unique aspects of poverty present and their solutions. On the other hand, the

regions experiencing the highest rates of poverty are Meranti Islands, with 22.98 percent, and Aceh Singkil at 19.15 percent, indicating that these locales require increased focus and efforts to enhance the economic situation of their residents (BPS, 2024). Poverty is defined as the inability of individuals to meet their essential nee

ds, such as food, clothing, housing, and healthcare. Nurkse's theory, known as The Vicious Circle of Poverty, posits that there are three primary contributors to poverty. The first factor is inadequate human resource development, which results in poor educational attainment; the second is imperfections in the market; and the third is a lack of capital, leading to lower productivity levels (Ayudia et al., 2024).

Education ranks among the socioeconomic elements that affect poverty levels. One way to assess whether the educational standards of a particular area or nation are adequate is by examining the average duration of schooling (Rafiqi, 2020). The unemployment rate serves as another key factor in understanding poverty. A high unemployment figure signifies deficiencies in job creation, which can aggravate the economic situation for the community and may jeopardize social stability. Therefore, it is essential to implement strategies to maintain a steady unemployment rate (Feriyandri & Maimunah, 2023). Lincolin asserts that there exists a strong correlation between elevated unemployment and poverty rates (Zaqiah et al., 2023). Per capita income, as calculated by GRDP per capita at current prices (ADHB), is a critical economic metric in poverty evaluation. The fluctuations in GRDP levels across different regions result from the region's capacity to effectively manage its local resources (Damanik & Sidauruk, 2020). Additionally, per capita spending acts as a vital indicator reflecting the community's capacity to meet essential needs. The number of individuals living in poverty is significantly shaped by the poverty threshold, which serves as a measure of poverty, since poor individuals are classified as those who have an average monthly per capita expenditure below this threshold (Rafiqi, 2020).

By employing the K-Means clustering approach, regions and municipalities on Sumatra Island will be categorized based on poverty statistics that include indicators such as the proportion of impoverished individuals, average schooling duration, unemployment percentage, income per individual, and spending per individual. This analysis aims to discern the distinctions in poverty rates across different clusters of regions and municipalities on Sumatra Island. Utilizing the most recent data from 2023, the findings of this study may serve as a basis for future poverty reduction initiatives. Furthermore, it is anticipated that the results will offer valuable guidance for policymakers to devise more suitable and impactful strategies aimed at fostering prosperity within the region.

METHOD

The information employed in this analysis is secondary data, consisting of cross-sectional data obtained from the Central Bureau of Statistics website, which represents information gathered simultaneously from numerous individuals. Secondary data refers to information that has been compiled and released by relevant organizations associated with this study. The indicators or variables examined in this research include the percentage of individuals living in poverty (PPM), average years of education (RRLS), unemployment rate

ISSN (e): 2829-7350 | ISSN(p): 2963-9441

(TPT), income per person (PDP), and expenditure per person (PNP) for the regencies/cities on Sumatra Island in the year 2023.

The research approach undertaken is descriptive quantitative, a methodology designed to outline the traits of a phenomenon or demographic using numerical values. This study gathers and scrutinizes numerical data to portray an accurate and objective representation of the analyzed scenario. By utilizing the K-Means Clustering analytical strategy, this study seeks to classify 154 regencies/cities in Sumatra Island into various clusters or groups to uncover poverty trends and the features contributing to poverty within each cluster. The findings of this research are intended to assist government officials in devising effective policies to combat poverty and enhance community welfare in the region.

Descriptive statistics is a method of analysis that encompasses gathering, handling, showcasing, and interpreting numerical or percentage data shown in tables or charts (Mayasari & Nugraha, 2023). Within descriptive statistics, information is conveyed by detailing or offering insights regarding the data.

Multicollinearity in relation to K-Means Clustering indicates a scenario where the factors applied for grouping data are closely related linearly. The examination for multicollinearity is conducted to assess how each variable or indicator correlates by analyzing the VIF (Variance Inflation Factor) figure. If the VIF number exceeds 10, it signifies the presence of multicollinearity, suggesting that the resulting estimation may be less precise (Hidayat, 2022).

In this research, the ideal quantity of clusters is identified by employing both the elbow and silhouette techniques. The elbow technique involves choosing the best number of clusters by identifying the point where the graph forms an elbow shape (Dito, 2020).

K-Means is a method that does not follow a hierarchical approach, indicating that it clusters data without depending on a structured order. This technique segments data into clusters according to shared attributes, ensuring that like data are grouped together while dissimilar data are placed in different clusters. The primary aim of data clustering is to reduce the designated objective function by diminishing the variability within each cluster and amplifying the variability across clusters (Mayasari & Nugraha, 2023).

According to Hair J et al (2009), the One Way Anova *F* statistic is utilized to assess whether a statistically significant distinction exists among the constructed clusters and their corresponding clustering variables (independent variables). ANOVA is frequently employed in studies that involve comparative assessments, aiming to analyze the dependent variable through comparisons among groups of observed independent samples (Hanifah, 2023). Hypothesis testing was performed to identify if notable differences exist in poverty traits among the established clusters. The study's hypotheses were as follows:

1. H0: There is no noteworthy difference in poverty traits among the poverty rate clusters of regencies/cities on Sumatera Island in 2023.
2. H1: There is a noteworthy difference in poverty traits among the poverty rate clusters of regencies/cities on Sumatera Island in 2023.

If the p-value is greater than 0.05, then the null hypothesis H0 is confirmed, while the alternative hypothesis H1 is dismissed. This indicates that accepting H0 suggests there is no noteworthy disparity in poverty traits among the distinct groups of regency or city poverty rates on Sumatra Island in the year 2023. Conversely, if the p-value is equal to or less than 0.05, the null hypothesis H0 is dismissed and the alternative hypothesis H1 is endorsed. This indicates that rejecting H0 implies there is a significant contrast in poverty characteristics across different poverty levels of districts or municipalities on Sumatra Island in 2023.

If the One Way Anova analysis indicates that noteworthy discrepancies exist in poverty attributes across various poverty level clusters or Municipal Districts on Sumatra Island, it becomes essential to perform a Post Hoc Difference Test. This statistical evaluation measures the significance of mean differences among groups after the overall differences have been established by ANOVA (Ostertagová & Ostertag, 2013). For this investigation, the Post Hoc difference test will apply Tukey testing, specifically the Tukey HSD (*Honestly Significant Difference*) test, which was created by John Tukey to pinpoint substantially differing indicators (Najih, 2024). Below are the hypotheses proposed for the Tukey HSD Post Hoc Test:

1. H0: No substantial differences exist in the poverty rate clusters when considering the poverty characteristics of regencies or cities on Sumatra Island in 2023.
2. H1: Substantial differences are present in the poverty level clusters regarding the poverty characteristics of regencies or cities on Sumatra Island in 2023.

If the p-value is greater than 0.05, we accept H0 and reject H1. This indicates that accepting H0 signifies there is no notable difference in poverty rate clusters according to the poverty traits of the regencies or cities on Sumatra Island in 2023. Conversely, if the p-value is less than or equal to 0.05, we reject H0 and accept H1. This implies that rejecting H0 indicates a significant difference in the clusters of poverty levels based on the poverty characteristics of the regencies or cities on Sumatra Island in 2023.

RESULTS AND DISCUSSION

Table 1. Results of Descriptive Statistical Analysis

	PPM (Percent)	RRLS (Year)	TPT (Percent)	PDP (Thousand Rupiah)	PNP (Thousand Rupiah)
Mean	10,017	9,003	4,498	67267	11230
Median	9,110	8,785	4,420	52255	10967
Maximum	22,980	13,040	10,860	391932	18990
Minimum	2,270	5,370	0,450	18420	6382

Source: *RStudio* Output

SINOMICS JOURNAL
International Journal of Social Science, Education, Communication and Economics
ISSN (e): 2829-7350 | ISSN(p): 2963-9441

The average rate of poverty stands at 10.017 percent, revealing that in the Sumatra regency or city, approximately 10% of individuals fall below the poverty threshold, translating to roughly 1 in every 10 individuals living in poverty. The mean duration of education among the population is recorded at 9.003 years, suggesting that individuals within the regency or city typically complete around 9 years of schooling. The average open unemployment rate is noted to be 4.498 percent, showing that the unemployment figures are relatively low. The average income per individual is 67,267 rupiah, signifying that the per capita earnings are considered to be quite substantial. The average per capita spending reaches 11,230 rupiah, illustrating that the typical spending per individual is considerably low in comparison to their income.

Table 2. Multicollinearity Test Results

PPM	RRLS	TPT	PDP	PNP
1,429849	1,705821	1,293916	1,259627	2,205811

Source: *RStudio* Output

According to the findings from the RStudio output mentioned earlier, the VIF value falls below 10, indicating an absence of multicollinearity in the data set. Consequently, it can be deduced that each variable such as PPM, RRLS, TPT, PDP, and PNP reflects non-multicollinearity, signifying that the indicators or variables do not exhibit a significant correlation with one another. This suggests that the model is free from multicollinearity, leading to more robust analysis outcomes and a clearer understanding of the results.

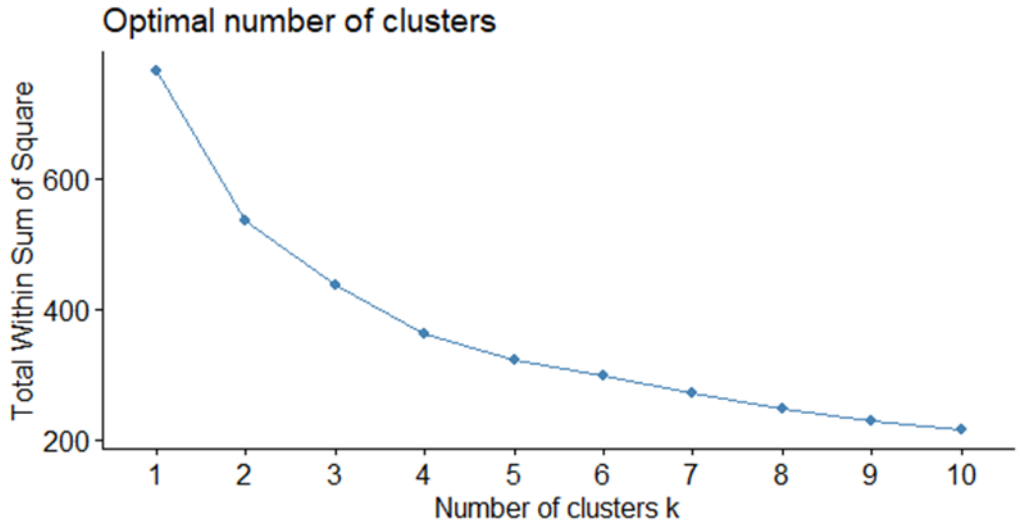

Figure 1. Graph of the Optimal Number of Clusters of the Elbow Method
Source: *RStudio* Output

Identifying the best number of clusters using the elbow method relies on finding the point where the graph has an elbow shape. As illustrated in Figure 1, it can be concluded that utilizing 2-3 clusters is sufficient since adding more clusters does not lead to a notable reduction in the total Within Sum of Squares. Thus, the ideal number of clusters to use is 3.

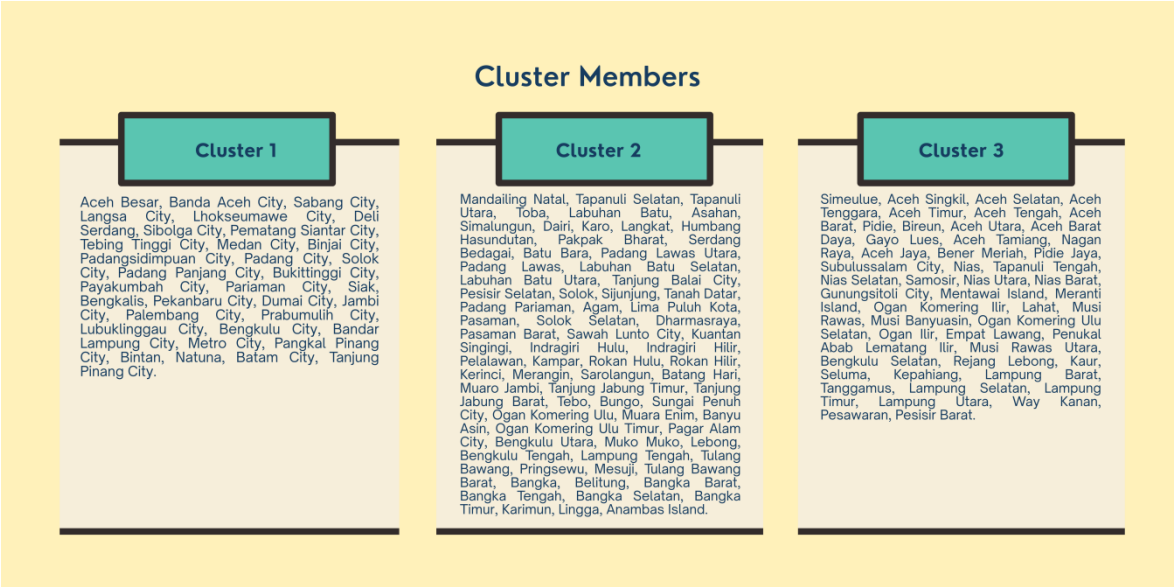

Figure 2. Cluster Result

According to the illustration in Figure 2, cluster 1 comprises five regencies and twenty-nine cities, featuring a predominance of developed regions that span across North Sumatra, South Sumatra, and segments of West Sumatra. Notable urban centers within this cluster include Medan City and Palembang City, alongside a variety of smaller municipalities. Cluster 2 is made up of sixty-seven regencies and four cities, representing a greater number of rural and isolated regions that largely encompass most of North Sumatra, Riau, Jambi, along with parts of West Sumatra and Lampung. In cluster 3, there are forty-seven regencies and two cities, which include a greater presence of remote locations and several minor islands, including Mentawai Island.

Table 3. Cluster Profilization

Indicator	Cluster 1	Cluster 2	Cluster 3
PPM	-0,58	-0,51	1,14
RRLS	1,39	-0,28	-0,57
TPT	1,01	-0,3	-0,27
PDP	0,53	0,09	-0,50
PNP	1,21	-0,02	-0,81

Source: *RStudio* Output

Referring to Table 3, the findings from the cluster profiling involving the PPM, RRLS, TPT, PDP, and PNP metrics indicate that the highest average value is highlighted in red, showing a classification of high, while average medium values are in blue, categorized as medium, and the low average values are marked in green, identified as low. The detailed description of each poverty cluster in the urban districts of Sumatra Island is presented below:

SINOMICS JOURNAL
International Journal of Social Science, Education, Communication and Economics
ISSN (e): 2829-7350 | ISSN(p): 2963-9441

a. Cluster 1 consists of regencies and cities that exhibit a lower average poverty level compared to clusters 2 and 3. This cluster is identified as a set of regencies or cities characterized by poverty, as evidenced by the metrics related to average years of education, unemployment rates, per capita income, and elevated consumer spending per capita.
b. Cluster 2 includes regencies and cities with a moderate average poverty level. This classification is recognized for its poverty characteristics based on average schooling years, open unemployment rates, per capita income, and average levels of per capita expenditure.
c. Cluster 3 identifies regencies and cities that show a high average poverty rate. This group encompasses regencies and cities with poverty traits that can be seen in the average years of education, existing unemployment rates, income per capita, and low per capita expenditure.

One Way Anova Test

Table 4. One Way Anova Test Results

Indicators for Each Cluster	*F*	Significant
PPM	118.113	0,000
RRLS	99.154	0,000
TPT	30.761	0,000
PDP	12.753	0,000
PNP	87.015	0,000

Source: *SPSS* Output

According to the findings in Table 4, the outcomes of the One Way Anova test for each variable indicate a considerable number with a p-value less than 0.05. As a result, the null hypothesis *H*0 is dismissed while the alternative hypothesis *H*1 is accepted, indicating a notable variation in poverty traits among the different poverty rate clusters present within the City Regencies of Sumatra Island for the year 2023. These variations are reflected in the distinctive poverty traits influenced by socioeconomic factors such as poverty, education levels, unemployment rates, income, and spending habits, which differ substantially among regencies and cities. Consequently, a Post Hoc Difference Test is feasible, and the results derived from the Post Hoc or Post-tests using the Tukey HSD technique can highlight the specific disparities between clusters, revealing which groups or clusters exhibit similarities and differences in PPM, RRLS, TPT, PDP, PNP.

Table 5. Post Hoc Difference Test Results

Cluster	Significant				
	PPM	**RRLS**	**TPT**	**PDP**	**PNP**
Cluster 1 and Cluster 2	0,858	0,000	0,000	0,068	0,000
Cluster 1 and Cluster 3	0,000	0,000	0,000	0,000	0,000

Cluster 2 and Cluster 3	0,000	0,052	0,984	0,002	0,000

Source: *SPSS* Output

Cluster 1 and cluster 2, characterized by low and medium poverty levels according to the PPM and PDP measures, exhibit no substantial differences. In contrast, the indicators RRLS, TPT, and PNP show considerable disparities. This implies that the regencies or cities located on Sumatra Island found in cluster 1 and cluster 2 differ in terms of RRLS, TPT, and PNP when defining poverty within the area. Cluster 1 and cluster 3, identified with low and high poverty levels through the PPM, RRLS, TPT, PDP, and PNP measures, display notable differences. This indicates that the characteristics of poverty according to these indicators in the Sumatra Island regencies or cities classified under cluster 1 and cluster 3 vary significantly from one another. Cluster 2 and cluster 3, which are associated with moderate and high poverty levels, show significant differences based on the RRLS, TPT, and PDP indicators. Additionally, there are notable differences in the PPM and PNP measures. This suggests that the regencies or cities on Sumatra Island categorized into cluster 2 and cluster 3 offer variations in PPM and PNP when defining poverty in the region.

According to the findings from the K-Means Clustering analysis, the poverty situation in regencies and cities across Sumatra Island in 2023 has been divided into three distinct groups, each representing varying levels of poverty. These groups include areas with low, medium, and high poverty rates. One of the Way Anova test outcomes indicates that the characteristics related to poverty among the districts and municipalities on Sumatra Island differ significantly among the three groups, focusing on metrics such as average educational attainment, open unemployment rate, income per person, and per capita expenditure. Given the unique characteristics of each group, it is clear that varying strategies or policies are necessary to address the poverty issues in each cluster. The Post Hoc Test using the Tukey HSD method offers a deeper insight into the poverty rate groups of regencies and cities on Sumatra Island:

1. Between Clusters 1 and 2, there are notable discrepancies in the RRLS, TPT, and PNP indicators, suggesting that the regencies and cities in these clusters require tailored interventions for the RRLS, TPT, and PNP metrics concerning poverty reduction.
2. Clusters 1 and 3 do not exhibit significant distinctions in any of the indicators, indicating that poverty in these two groups can be approached uniformly by implementing similar policies across the regions.
3. Clusters 2 and 3 show significant differences in the RRLS, TPT, and PDP indicators, implying that poverty management for regencies and cities in these clusters must adopt distinct strategies regarding the RRLS, TPT, and PDP metrics to effectively alleviate poverty.

CONCLUSION

Poverty in the regions and cities of Sumatra Island in 2023 shows varying degrees of severity when analyzed through K-Means Clustering. The first cluster includes 34 regencies and cities with a low incidence of poverty, marked by high average years of education,

ISSN (e): 2829-7350 | ISSN(p): 2963-9441

elevated open unemployment rates, substantial per capita income, and significant per capita spending. The second cluster comprises 71 regencies and cities with a moderate level of poverty, exhibiting average years of education, a medium open unemployment rate, average per capita income, and mid-range per capita expenditures. The third cluster contains 49 regencies and cities that display low average years of education, minimal open unemployment, reduced per capita income, and low per capita expenditure.

Clusters one, two, and three reveal notable variations in poverty levels when assessed through indicators like the percentage of impoverished individuals, average years of education, rates of open unemployment, per capita income, and per capita expenditure. The results from the clustering can illuminate the poverty characteristics inherent to each regency and city and are anticipated to guide government strategies in addressing poverty effectively while enhancing welfare in regions facing significant poverty challenges. Cluster three showcases the lowest poverty level in comparison to clusters one and two, suggesting that the city districts in cluster three may serve as focal points for government initiatives aimed at alleviating poverty and fostering community welfare.

REFERENCES

Amalia, F., & Emalia, Z. (2022). Fenomena Kelimpahan Sumber Daya Alam dan Natural Resource Curse Dalam Perspektif Ekonomi Di Pulau Sumatera. *BULLET : Jurnal Multidisiplin Ilmu*, *01*(5), 737–750.

Ayudia, N., Ciptawaty, U., Wahyudi, H., Yuliawan, D., & Ratih, A. (2024). Faktor-Faktor yang Mempengaruhi Tingkat Kemiskinan pada Daerah Tertinggal di Pulau Sumatera Berdasarkan Tipologi Klassen. *Journal on Education*, *06*(03), 17112–17121.

BPS. (2024). *Persentase Penduduk Miskin (PO) Menurut Kabupaten/Kota (Persen), 2022-2024*. Badan Pusat Statistik. https://www.bps.go.id/id/statistics-table/2/NjIxIzI%253D/persentase-penduduk-miskin--p0--menurut-kabupaten-kota.html

Damanik, R. K., & Sidauruk, S. A. (2020). Pengaruh Jumlah Penduduk Dan Pdrb Terhadap Kemiskinan Di Provinsi Sumatera Utara. *Jurnal Darma Agung*, *28*(3), 358. https://doi.org/10.46930/ojsuda.v28i3.800

Dito, B. S. dan G. A. (2020). *KMeans*. RPubs By RStudio. https://rpubs.com/bagusco/kmeans

Feriyandri, P. D., & Maimunah, E. (2023). Pengaruh Angkatan Kerja dan Investasi terhadap Produk Domestik Regional Bruto di Provinsi Lampung. *Journal on Education*, *6*(1), 8122–8133. https://doi.org/10.31004/joe.v6i1.4230

Hair J, R, A., Babin B, & Black W. (2009). Multivariate Data Analysis (Seven Ed). In *Pearson: Vol. 7 edition* (p. 761).

Hanifah, F. A. (2023). *Penggunaan Analisis One-Way ANOVA Pada Kasus Pengujian Pertumbuhan Produksi Maggot Melalui Kombinasi Sampah Rumah Tangga dan Daun Kering*. RPubs By RStudio. https://rpubs.com/fitriaamalia/miniprojectkomstat

Hidayat, A. (2022). *Klasterisasi Penyebaran Covid-19 di Indonesia Berdasarkan Provinsi Menggunakan K-Means Cluster*. RPubs By RStudio. https://rpubs.com/Anoe/lbb-

kmeans

Mayasari, S. N., & Nugraha, J. (2023). Implementasi K-Means Cluster Analysis untuk Mengelompokkan Kabupaten/Kota Berdasarkan Data Kemiskinan di Provinsi Jawa Tengah Tahun 2022. *KONSTELASI: Konvergensi Teknologi Dan Sistem Informasi*, *3*(2), 317–329. https://doi.org/10.24002/konstelasi.v3i2.7200

Najih. (2024). Uji Tukey/ Honest Significantly Difference/ Uji Beda Nyata Jujur. *RPubs By RStudio*. https://rpubs.com/najih/UjiTukey

Ostertagová, E., & Ostertag, O. (2013). Methodology and Application of Oneway ANOVA. *American Journal of Mechanical Engineering*, *1*(7), 256–261. https://doi.org/10.12691/ajme-1-7-21

Rafiqi, A. (2020). Pengaruh Rata-rata Lama Sekolah, Pengeluaran Riil Perkapita, Pertumbuhan Ekonomi dan Pengangguran Terhadap Tingkat Kemiskinan di Provinsi D.I Yogyakarta. *Skripsi*.

Salsabila, R. (2020). Pengaruh Kemiskinan dan Pengeluaran Pemerintah Dalam Sektor Pendidikan Terhadap Indeks Pembangunan Manusia di Wilayah Sumatera. *Skripsi*, *7*(2).

Zaqiah, A., Triani, M., & Yeni, I. (2023). Pengaruh Pendidikan, Pengangguran dan Jumlah Penduduk Terhadap Tingkat Kemiskinan di Indonesia. *Jurnal Kajian Ekonomi Dan Pembangunan*, *5*(3), 33. https://doi.org/10.24036/jkep.v5i3.15284