

# Clustering of Regencies/Cities Based on Factors Influencing Poverty in West Sumatra Using K-Medoids

Afifah Hardi<sup>1</sup>, Dony Permana<sup>1\*</sup>, dan Denny Armelia<sup>2</sup>

<sup>1</sup>Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

<sup>2</sup>BPS Provinsi Sumatera Barat, Padang, Indonesia

\*Corresponding author: donypermana@fmipa.unp.ac.id

Submitted : 07 Juni 2025

Revised : 25 November 2025

Accepted : 01 Desember 2025

## ABSTRACT

Poverty remains a significant issue in Indonesia, particularly in West Sumatra Province, where regional disparities persist despite a national decline in poverty rates. This study aims to classify the 19 regencies/cities in West Sumatra based on key socioeconomic indicators to support more targeted and effective poverty alleviation policies. Using a quantitative descriptive approach, the research applies the K-Medoids clustering method to group regions according to four indicators: Gross Regional Domestic Product (GRDP) per capita, Human Development Index (HDI), Open Unemployment Rate (OUR), and Gini Ratio. Secondary data for the year 2024 were obtained from the official website of the Central Bureau of Statistics of West Sumatra. Prior to clustering, data standardization using Z-score transformation was performed, and multicollinearity was tested using the Variance Inflation Factor (VIF). The silhouette method indicated that the optimal number of clusters is four. The clustering analysis revealed four distinct groups: (1) underdeveloped areas with low income and human development but high inequality; (2) moderately developed areas with stable unemployment and low income inequality; (3) urbanized areas with high income and human development but also high unemployment and inequality; and (4) a single metropolitan area with high economic and human development and moderate inequality. The findings highlight the importance of region-specific strategies in addressing poverty, considering the diverse economic and social conditions across regions. The results can serve as a basis for designing equitable and effective socioeconomic development policies.

**Keywords:** Clustering, Gini Ratio, Human Development Index, K-Medoids, Poverty.



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

## I. PENDAHULUAN

Kemiskinan adalah masalah yang masih sangat diperhatikan di Indonesia, termasuk di Sumatera Barat. Walaupun data dari Badan Pusat Statistik pada tahun 2023 menunjukkan bahwa angka kemiskinan secara nasional dan regional mengalami penurunan, perbedaan antar daerah tetap terasa. Pada bulan September 2024, jumlah orang miskin di Sumatera Barat tercatat sebanyak 315,43 ribu orang, berkurang dari 349,53 ribu orang pada tahun 2015. Namun, ketidakmerataan masih tinggi, seperti yang terlihat di Kabupaten Kepulauan Mentawai yang memiliki tingkat kemiskinan tertinggi di provinsi ini, yaitu 13,89 persen menurut BPS Sumbar pada tahun 2024 (BPS Sumbar, 2024).

Penyebab kemiskinan di berbagai kabupaten/kota biasanya berbeda-beda. Karena itu, sangat penting untuk mengelompokkan daerah berdasarkan ciri-ciri ekonomi dan sosial tertentu agar kebijakan untuk mengatasi kemiskinan bisa lebih efektif. Beberapa indikator yang relevan untuk analisis ini meliputi Produk Domestik Regional Bruto (PDRB) per orang, Indeks Pembangunan Manusia (IPM), Tingkat Pengangguran Terbuka (TPT), dan Rasio Gini.

Produksi Domestik Regional Bruto (PDRB) per orang menunjukkan seberapa kuat ekonomi suatu daerah dalam memproduksi barang dan jasa yang menjadi sumber utama penghasilan masyarakat. Ketika PDRB per kapita lebih tinggi, biasanya daya beli dan kesejahteraan ekonomi penduduk di daerah itu juga meningkat. Namun, angka ini tidak selalu menunjukkan bahwa distribusinya adil di antara berbagai kelompok masyarakat, terutama jika perhitungannya masih menggunakan harga yang berlaku dan belum memperhitungkan inflasi (Alhudori, 2017).

Di sisi lain, Indeks Pembangunan Manusia (IPM) menunjukkan kualitas sumber daya manusia dengan tiga aspek utama, yaitu panjang umur dan kesehatan, pendidikan, serta kualitas hidup yang baik. IPM menjadi indikator penting

karena tidak hanya mencerminkan kemajuan dalam kesehatan dan pendidikan, tetapi juga menggambarkan kemampuan seseorang untuk menjalani kehidupan yang produktif dan berarti (Dewi *et al.*, 2017).

Berdasarkan penelitian (Yuniarti & Imaningsih, 2022), Tingkat Pengangguran Terbuka (TPT) mengindikasikan seberapa besar bagian dari angkatan kerja yang tidak memiliki pekerjaan. Tingginya angka pengangguran sering kali mencerminkan bahwa pasar tenaga kerja tidak berfungsi secara optimal dan dapat menjadi tanda bahwa ekonomi tidak mampu menyerap jumlah orang yang ingin bekerja. Situasi ini dapat membuat orang lebih rentan terhadap kemiskinan, terutama bagi masyarakat yang berpenghasilan tidak tetap di sektor informal.

Sementara itu, Gini Ratio berfungsi untuk menilai ketidakmerataan distribusi pendapatan antarpenduduk di suatu daerah. Jika nilai Gini Ratio hampir mencapai 1, itu menunjukkan ketidakmerataan yang signifikan, yang dapat menghambat usaha untuk mengatasi kemiskinan karena hanya sekelompok kecil orang yang merasakan keuntungan dari pembangunan. Data mengungkapkan bahwa pada September 2024, nilai Gini Ratio di Sumatera Barat adalah 0,287, naik dari 0,283 pada Maret 2024.

Penelitian ini menerapkan teknik pengelompokan *K-Medoids* untuk mengategorikan daerah-daerah di Provinsi Sumatera Barat menurut indikator yang ada. Proses analisis menggunakan data dari tahun 2024. Penelitian ini bertujuan untuk menemukan pola pengelompokan wilayah yang didasarkan pada karakteristik kemiskinan, sehingga temuan ini bisa menjadi acuan dalam menyusun kebijakan pembangunan sosial ekonomi yang lebih adil dan efektif.

## II. METODE PENELITIAN

### A. Jenis Penelitian dan Sumber Data

Penelitian ini mengimplementasikan teknik kuantitatif dengan pendekatan deskriptif dan analisis kluster, dan menggunakan metode *K-Medoids* untuk mengelompokkan Kabupaten/Kota di Provinsi Sumatera Barat yang dipengaruhi oleh berbagai faktor kemiskinan. Sumber data untuk penelitian ini bersifat sekunder, diambil dari situs resmi BPS Provinsi Sumatera Barat. Dalam analisis ini, penulis menggunakan data mengenai Produk Domestik Regional Bruto per Kapita (X1), Indeks Pembangunan Manusia (X2), Tingkat Pengangguran Terbuka (X3), dan Gini Ratio (X4) untuk setiap daerah di Provinsi Sumatera Barat tahun 2024.

### B. Teknik Analisis Data

#### 1. Pengecekan *Outlier*

Pengecekan *outlier* dilakukan menggunakan salah satu metode pengecekan *outlier* yaitu metode *Quan* (*Quantile-based method*). Metode tersebut merupakan metode yang bertujuan untuk menganalisis distribusi data berdasarkan kuantil dengan membagi data menjadi beberapa bagian yang sama besar, kemudian dengan metode *Quan* akan mengidentifikasi data yang menyimpang jauh dari data yang lain atau disebut dengan nilai ekstrim. Namun demikian, didalam analisis *K-Medoids* nantinya jika terdapat *outlier* pada data, data *outlier* tetap digunakan dalam analisis pembentukan *cluster* dikarenakan metode tersebut dirancang untuk mengatasi data yang mengandung *outlier*, metode *K-Medoids* memastikan bahwa hasil pengelompokan tetap representative terhadap keseluruhan data walaupun terdapat *outlier* (Sulistiyawati, 2024).

#### 2. Pengecekan Multikolinearitas

Multikolinearitas merupakan keadaan di mana variabel independen saling berkaitan secara erat. Oleh karena itu, variabel-variabel ini sebaiknya tidak mengalami multikolinearitas. Salah satu cara untuk mengetahui adanya multikolinearitas adalah dengan menghitung nilai tolerance atau VIF (*Variance Inflation Factor*). Apabila nilai VIF berada di bawah 0,1 atau di atas 10, maka dapat disimpulkan bahwa multikolinearitas terjadi. (Safitri *et al.*, 2021).

$$VIF = \frac{1}{1 - r^2} \quad (1)$$

$$r_{x_1x_2} = \frac{(n \sum x_1x_2 - (\sum x_1) \cdot \sum x_2)}{\sqrt{n \sum x_1^2 - (\sum x_1)^2} \cdot \sqrt{n \sum x_2^2 - (\sum x_2)^2}}$$

Keterangan:

n = Jumlah data yang digunakan

$r_{x_1x_2}$  = Koefisien Determinasi

#### 3. Teknik *Clustering*

*Clustering* adalah salah satu metode dalam data mining yang berfungsi untuk mengatur data yang sebelumnya tidak terklasifikasi atau tidak memiliki kategori yang jelas. Dalam algoritma

pengelompokan, data dibagi menjadi beberapa kelompok berdasarkan kesamaan di antara data tersebut. Proses ini dilakukan tanpa acuan atau label tertentu untuk menentukan bagaimana pengelompokan dilakukan. Secara umum, terdapat dua tipe pendekatan dalam pengelompokan, yaitu:

- *Partitional Clustering*: metode pengelompokan di mana setiap objek hanya termasuk dalam satu cluster yang terpisah sepenuhnya dari *cluster* lainnya.
- *Hierarchical Clustering*: metode pengelompokan yang menyusun data dalam struktur pohon bertingkat (*hierarki*), di mana *cluster-cluster* tersusun dalam beberapa level (Pradipta, 2021).

a. *K-Medoids*

Metode *K-Medoids*, yang sering disebut sebagai *Partitioning Around Medoids* (PAM), merupakan teknik pengelompokan yang memanfaatkan objek data sebagai pusat untuk tiap cluster, yang dikenal sebagai medoid. Tujuan dari metode ini adalah untuk mengurangi total perbedaan antara setiap titik data dalam cluster dengan medoid yang dipilih sebagai pusatnya. Perbedaan utama yang membedakan *K-Medoids* dari *K-Means* adalah cara menentukan pusat cluster: *K-Means* menggunakan rata-rata nilai dari data dalam cluster sebagai pusat, sedangkan *K-Medoids* memilih satu objek data sebagai wakil pusat cluster. Salah satu keuntungan dari algoritma *K-Medoids* adalah kemampuannya untuk mengatasi kelemahan *K-Means*, terutama dalam hal sensitivitas terhadap kebisingan dan *outlier*, yang dapat menyebabkan ketidaksesuaian objek sehingga jauh dari pola distribusi data (Haumahu & Matdoan, 2023).

b. Penentuan Jumlah *Cluster*

Penentuan jumlah *cluster* digunakan dengan menggunakan metode *silhoutette*. Metode ini menggunakan koefisien siluet yang menggabungkan pemisah dan kohesi. Semakin besar koefisien *silhoutette*, semakin baik klaster tersebut (Putri & Betha Nurina, 2023).

$$SC = \max_k SI(k) \quad (2)$$

Keterangan:

$SC$  = *Silhouette Coefficient*

$SI$  = *Silhouette Index Global*

$k$  = jumlah *cluster*

c. Pengelompokan *Cluster*

Pengelompokan dilakukan setelah mendapatkan jumlah *cluster* optimum yang ditentukan menggunakan metode *silhoutette*. Algoritma *K-Medoids* kemudian mengelompokkannya berdasarkan kemiripan karakteristik indikator terhadap medoid.

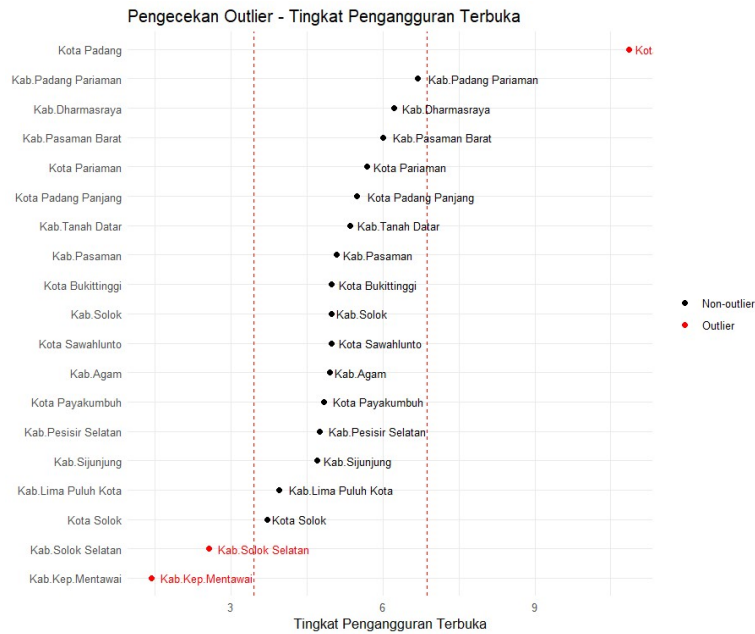
4. Profilisasi dan Visualisasi

Profilisasi digunakan untuk menggambarkan karakteristik masing-masing *cluster* dengan menghitung rata-rata dari setiap variabel dalam tiap *cluster*, ini digunakan untuk memberikan interpretasi mengenai perbedaan pola antar *cluster*. Visualisasi hasil *clustering* dilakukan untuk mempermudah interpretasi hasil *cluster*, baik dalam bentuk heatmap maupun peta tematik.

### III. HASIL DAN PEMBAHASAN

1. Pengecekan *Outlier*

Analisis *K-Medoids cluster* digunakan untuk mengelompokkan data yang mengandung *outlier*, oleh karena itu dilakukan pengecekan *outlier* dengan metode quan sebagai berikut.



**Gambar 1.** Pengecekan Outlier

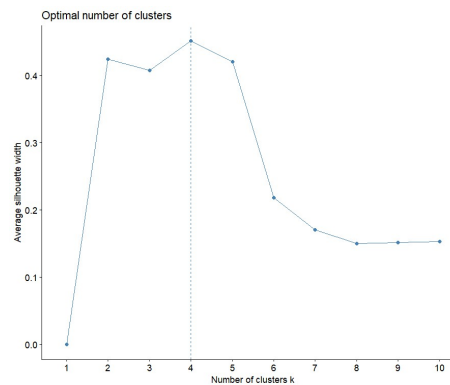
Dari *output* yang ada pada Gambar 3 dapat dilihat bahwa dengan menggunakan metode quan ditemukan 3 data *outlier* dan 16 data *non-outlier*. Data *outlier* dilambangkan dengan titik warna merah, sedangkan data bukan *outlier* dilambangkan dengan titik berwarna hitam dan 3 data *outlier* tersebut terlibat pada data Kota Padang, Kabupaten Solok Selatan, Kabupaten Kepulauan Mentawai.

2. Uji multikolinearitas

**Tabel 1.** Nilai VIF Masing-Masing Variabel

Nilai VIF	Variabel
1,00	PDRB Per Kapita
1,76	IPM
2,30	TPT
3,58	Gini Ratio

3. Analisis *K-Medoids Clustering*  
a. Menentukan Jumlah *Cluster*

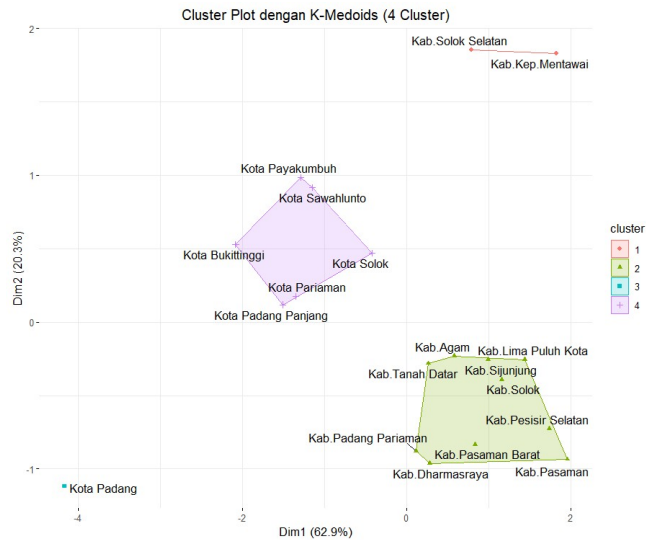


**Gambar 2.** Plot *Silhouette*

Dengan menggunakan metode *Silhouette* dapat memperkirakan jumlah *cluster* yang terbentuk. Dari output yang ada pada Gambar 4 terdapat saran dari metode *Silhouette* melalui garis putus-putus yang menunjukkan pada angka 4 sebagai *cluster* optimum. Metode *Silhouette* menggunakan nilai rata-rata jika semakin tinggi nilai semakin baik klasterisasi.

b. Pengelompokan *cluster*

Melalui analisis cluster yang menggunakan algoritma K-Medoids, diperoleh hasil berupa grafik *cluster* yang terdiri atas 4 *cluster*. Dalam pembuatan grafik kluster, jarak yang dipakai adalah jarak *Euclidean*. Ini karena jenis jarak *Euclidean* sesuai untuk data yang tidak mengalami multikolinearitas.



Gambar 3. Output K-Medoids Cluster

Dari hasil yang ditunjukkan pada Gambar 5, terlihat adanya 4 kelompok yang terbentuk berdasarkan perhitungan kesamaan antara indikator kemiskinan di Provinsi Sumatera Barat. Dalam grafik hasil kluster, terdapat 4 warna yang berbeda: merah untuk kelompok 1, hijau untuk kelompok 2, biru untuk kelompok 3, dan ungu untuk kelompok 4. Masing-masing warna memiliki karakteristik yang unik.

Pada output yang ditampilkan di Gambar 5, kita dapat mengamati anggota-anggota di setiap kelompok. Berikut adalah daftar anggota yang terdapat dalam setiap kelompok.

Tabel 2. Hasil Pengelompokan Cluster

Cluster	Jumlah	Anggota cluster
1	2	Kab.Solok Selatan, Kab.Kepulauan Mentawai
2	10	Kab.Agam, Kab.Tanah Datar, Kab.Lima Puluh Kota, Kab.Sijunjung, Kab.Solok, Kab.Pesisir Selatan, Kab.Padang Pariaman, Kab.Pasaman Barat, Kab.Dharmasraya, Kab.Pasaman
3	6	Kota Payakumbuh, Kota Bukittinggi, Kota Sawahlunto, Kota Solok, Kota Pariaman, Kota Padang Panjang
4	1	Kota Padang

4. Profilisasi dan visualisasi

Setelah diketahui jumlah serta anggota dari cluster, langkah berikutnya adalah melakukan profiling untuk memahami karakteristik setiap kelompok. Dalam proses profiling ini, digunakan rata-rata variabel yang ada pada setiap cluster.

**Tabel 3.** Hasil Profilisasi

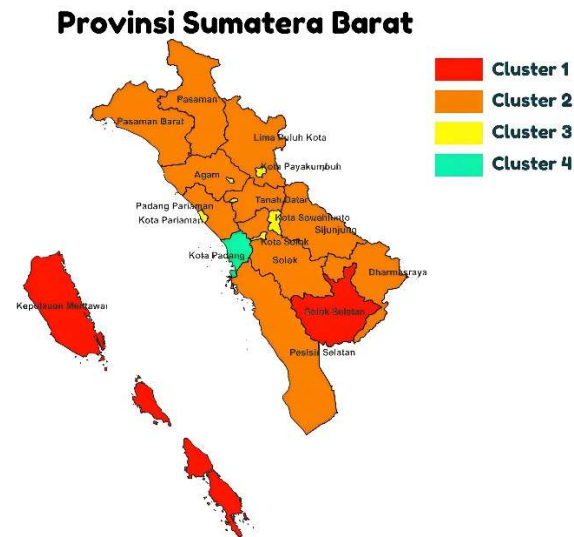
<i>cluster</i>	X1	X2	X3	X4
1	-0,51	-1,29	-1,68	0,89
2	-0,64	-0,50	0,08	-0,73
3	1,72	1,87	3,10	1,95
4	0,95	0,95	0,09	0,60

Berdasarkan Tabel 4, maka dapat dilakukan interpretasi sebagai berikut:

- a. *Cluster 1* merupakan kelompok kabupaten/kota dengan pendapatan per kapita yang rendah, kualitas pembangunan manusia yang sangat rendah, tingkat pengangguran yang sangat rendah, serta tingkat ketimpangan pendapatan yang tinggi. Artinya, Kabupaten/Kota-Kabupaten/Kota dalam klaster ini masih tertinggal secara ekonomi dan kesejahteraan, namun memiliki tingkat pengangguran yang rendah. Meski begitu, kesenjangan antara yang kaya dan miskin masih cukup besar.
- b. *Cluster 2* merupakan kelompok Kabupaten/Kota pendapatan per kapita yang rendah, pembangunan manusia yang masih tergolong rendah, tingkat pengangguran yang relatif stabil, serta ketimpangan pendapatan yang rendah. Hal ini menunjukkan bahwa Kabupaten/Kota dalam klaster ini tidak tergolong kaya atau maju, tetapi memiliki kondisi sosial yang cukup seimbang. Pengangguran tidak terlalu tinggi, dan distribusi penghasilan cenderung merata. Klaster ini bisa disebut sebagai wilayah dengan kondisi sedang dan cukup stabil.
- c. *Cluster 3* merupakan kelompok Kabupaten/Kota yang memiliki pendapatan per kapita dan pembangunan manusia yang sangat tinggi, namun juga diikuti dengan tingkat pengangguran dan ketimpangan pendapatan yang sangat tinggi. Ini menunjukkan bahwa Kabupaten/Kota-Kabupaten/Kota ini sudah maju secara ekonomi dan pendidikan, tetapi menghadapi masalah sosial yang serius. Banyak masyarakat yang belum mendapatkan manfaat dari kemajuan tersebut, sehingga pengangguran dan kesenjangan masih tinggi. Wilayah dalam klaster ini kemungkinan besar adalah kota besar atau pusat ekonomi yang belum inklusif.
- d. *Cluster 4* merupakan kelompok Kabupaten/Kota dengan pendapatan per kapita dan pembangunan manusia yang tinggi, tingkat pengangguran yang stabil, serta tingkat ketimpangan pendapatan yang cukup tinggi namun masih dalam batas wajar. Kabupaten/Kota dalam klaster ini menunjukkan perkembangan yang baik, baik dari sisi ekonomi maupun kualitas hidup masyarakat. Pengangguran cukup terkendali, meskipun masih ada tantangan dalam hal pemerataan pendapatan.

Visualisasi *K-Medoids Clustering*

Untuk mempermudah dalam melihat hasil *clustering*, maka dilakukan visualisasi dari hasil *cluster* berupa sebuah peta menggunakan *software* QGIS.



Gambar 4. Visualisasi Clustering

Dari Gambar 6, terlihat bahwa di *cluster* 1 ada 2 Kabupaten, di *cluster* 2 ada 10 Kabupaten/Kota, di *cluster* 3 ada 6 Kota, dan di *cluster* 4 ada 1 Kota yang terdapat di Sumatera Barat. Masing-masing *cluster* ini memiliki faktor-faktor yang berbeda yang menyebabkan kemiskinan.

#### IV. KESIMPULAN

Penelitian ini berhasil mengelompokkan 19 kabupaten dan kota di Provinsi Sumatera Barat dengan menggunakan metode *K-Medoids*. Pengelompokan ini didasarkan pada empat indikator utama kemiskinan, yaitu PDRB per kapita, Indeks Pembangunan Manusia (IPM), Tingkat Pengangguran Terbuka (TPT), dan Gini Ratio. Analisis yang dilakukan menghasilkan empat cluster, masing-masing dengan karakteristik sosial ekonomi yang berbeda. *Cluster* 1 terdiri dari daerah dengan pendapatan dan Pembangunan manusia rendah, namun mereka memiliki tingkat pengangguran yang rendah. *Cluster* 2 menunjukkan daerah dengan kondisi ekonomi menengah dan distribusi pendapatan yang merata. *Cluster* 3 adalah wilayah urban yang maju, namun memiliki tingkat pengangguran dan ketimpangan yang tinggi. Sementara itu, *Cluster* 4 hanya meliputi Kota Padang yang menunjukkan tingkat pendapatan dan pembangunan yang tinggi, meskipun ketimpangannya tetap dalam batas wajar.

#### DAFTAR PUSTAKA

- Alhudori, M. (2017). Pengaruh IPM, PDRB dan Jumlah Pengangguran Terhadap Penduduk Miskin di Provinsi Jambi. *Ekonomis : Jurnal of Economics and Business*, 1(1), 113–124.
- BPS Sumbar. (2024). Provinsi Sumatera Barat dalam angka 2024. *Badan Statistik Provinsi Sumatera Barat*, 54, 282–283.
- Dewi, N., Yusuf, Y., & Iyan, R. Y. (2017). Pengaruh Kemiskinan Dan Pertumbuhan Ekonomi Terhadap Indeks Pembangunan Manusia Di Provinsi Riau. *JOM Fekon*, 4(1), 870–882.
- Haumahu, G., & Matdoan, M. Y. (2023). Algoritma *K-Medoids Clustering* Untuk Mengelompokkan Tingkat Kemiskinan Pada Kabupaten Dan Kota Di Kepulauan Maluku Dan Papua. *VARIANCE: Journal of Statistics and Its Applications*, 4(2), 81–88. <https://doi.org/10.30598/variancevol4iss2page81-87>
- Pradipta, G. A. (2021). Modul Ajar Data Mining.
- Putri, V., & Betha Nurina, S. (2023). Perbandingan Metode *Elbow* dan *Silhouette* untuk Penentuan Jumlah Kluster yang Optimal pada *Clustering* Produksi Padi menggunakan Algoritma *K-Means*. 9(November), 547–558.
- Safitri, P. N., Aristawidya, R., & Faradilla, S. B. (2021). Klasterisasi Faktor-Faktor Kemiskinan Di Provinsi Jawa Barat Menggunakan *K-Medoids Clustering*. *Journal of Mathematics Education and Science*, 4(2), 75–80. <https://doi.org/10.32665/james.v4i2.242>
- Sulistiyawati, Y. (2024). Implementasi *K-Medoids Clustering* Dalam Pengelompokan Provinsi Di Indonesia Berdasarkan Kepesertaan Aktif Jaminan Sosial Tenaga Kerja. *Emerging Statistics and Data Science Journal*, 2(3), 368–381. <https://doi.org/10.20885/esds.vol2.iss.3.art28>
- Yuniarti, Q., & Imaningsih, N. (2022). Pengaruh Pertumbuhan Ekonomi, Tingkat Kemiskinan dan Indeks

Pembangunan Manusia terhadap Tingkat Pengangguran Terbuka di Kabupaten Sidoarjo. *Ekonomis: Journal of Economics and Business*, 6(1), 44. <https://doi.org/10.33087/ekonomis.v6i1.474>