

MARKOV CHAIN K-MEANS CLUSTER MODEL DALAM DINAMIKA TRANSISI TINGKAT KEMISKINAN BERDASARKAN INDEKS KEDALAMAN DAN KEPARAHAN KEMISKINAN DI PROVINSI SUMATERA UTARA

Desto Wardana Siregar

Matematika, FMIPA, Universitas Negeri Medan, Kota Medan, Indonesia

wardanadesto@gmail.com

Pardomuan Sitompul

Matematika, FMIPA, Universitas Negeri Medan, Kota Medan, Indonesia

ptmath@unimed.ac.id

* wardanadesto@gmail.com

Abstrak

Kemiskinan merupakan masalah multidimensi yang kompleks dan memerlukan pendekatan analitis untuk memahami dinamikanya. Penelitian ini bertujuan untuk menganalisis dinamika transisi tingkat kemiskinan di Provinsi Sumatera Utara yang diukur berdasarkan Indeks Kedalaman Kemiskinan (*Poverty Gap Index-P₁*) dan Indeks Keparahan Kemiskinan (*Poverty Severity Index-P₂*) menggunakan *Markov Chain K-Means Cluster Model*. Data yang digunakan adalah data sekunder dari Badan Pusat Statistik Sumatera Utara tahun 2023-2024. Penelitian ini diuraikan ke dalam dua bagian utama yakni *K-Means Clustering* untuk segmentasi kabupaten/kota berdasarkan tingkat kemiskinan, dan analisis *Markov Chain* untuk memodelkan transisi antar *cluster* kemiskinan. Hasil klasterisasi menunjukkan bahwa data terbagi menjadi tiga *cluster* optimal: *Cluster 1* dengan *centroid* (0,02251, 0,00561) pada tahun 2023 dan (0,01885, 0,0037) pada tahun 2024, *Cluster 2* dengan *centroid* (0,01384, 0,003263) pada tahun 2023 dan (0,01225, 0,00275) pada tahun 2024, dan *Cluster 3* dengan *centroid* (0,00833, 0,001593) pada tahun 2023 dan (0,007778, 0,001328) pada tahun 2024. Analisis *Markov Chain* menghasilkan matriks peluang transisi yang memenuhi sifat *ergodic* (*irreducible, aperiodic dan positive recurrent*) dengan kondisi *steady state* dicapai pada periode ke-18 dimana dalam model ini ditemukan bahwa pada tahun 2041, Provinsi Sumatera Utara diproyeksikan sepenuhnya terbebas dari *cluster* kemiskinan tinggi (ekstrem) dengan probabilitas sebesar 76,69%.

Kata Kunci: kemiskinan, *clustering*, *Markov Chain*, *ergodic*, *steady state*.

Abstract

Poverty is a complex multidimensional problem that requires an analytical approach to understand its dynamics. This study aims to analyze the transition dynamics of poverty levels in North Sumatra Province, measured by the Poverty Gap Index (P_1) and the Poverty Severity Index (P_2), using the Markov Chain K-Means Cluster Model. The data used are secondary data from the North Sumatra Central Bureau of Statistics for the years 2023–2024. The study is divided into two main parts: K-Means Clustering for the segmentation of regencies/cities based on poverty levels, and Markov Chain analysis to model transitions between poverty clusters. The clustering results show that the data is optimally divided into three clusters: Cluster 1 with centroids (0.02251, 0.00561) in 2023 and (0.01885, 0.0037) in 2024; Cluster 2 with centroids (0.01384, 0.003263) in 2023 and (0.01225, 0.00275) in 2024; and Cluster 3 with centroids (0.00833, 0.001593) in 2023 and (0.007778, 0.001328) in 2024. The Markov Chain analysis generates a transition probability matrix satisfying ergodic properties (irreducible, aperiodic, and recurrent), achieving steady state conditions at the 18th period. The model projects that by 2041, North Sumatra Province has a 76.69% probability of complete transition out of the high-extreme poverty cluster.

Keywords: poverty, clustering, Markov Chain, ergodic, steady state.

PENDAHULUAN

Kemiskinan merupakan permasalahan sosial yang kompleks dan multidimensi yang dihadapi oleh hampir seluruh negara di dunia, termasuk

Indonesia (Amelia et al., 2025). Kemiskinan adalah keadaan dimana seseorang atau sekelompok orang, baik laki-laki maupun perempuan, tidak mampu memenuhi hak-hak dasar mereka seperti pangan, pendidikan, kesehatan, perumahan, air bersih,

transportasi dan sanitasi (Badan Pusat Statistik, 2024a). Sedangkan menurut *United Nations* kemiskinan mencakup rendahnya pendapatan dan sumber daya produktif untuk memastikan kehidupan yang berkelanjutan, kekurangan gizi dan kelaparan, kesehatan yang buruk, sulitnya mengakses pendidikan dan layanan dasar, peningkatan penyakit yang menyebabkan morbiditas dan mortalitas, tunawisma dan perumahan yang tidak memadai, diskriminasi sosial dan lingkungan yang tidak aman, kurangnya partisipasi dalam pengambilan keputusan dan dalam sipil, sosial dan kehidupan budaya (United Nation, 1995).

Berdasarkan cara pengukurannya, jenis kemiskinan yang sering digunakan di Indonesia bahkan oleh lembaga pemerintahan seperti Badan Pusat Statistik karna bersifat universal adalah kemiskinan absolut (Adon & Jeraman, 2023), dimana penduduk miskin didefinisikan sebagai penduduk dengan rata-rata pengeluaran per kapita per bulan di bawah garis kemiskinan. Garis kemiskinan atau batas kemiskinan adalah tingkat minimum pendapatan yang dianggap perlu dipenuhi untuk memperoleh standar hidup yang mencukupi di suatu wilayah dengan Garis Kemiskinan (GK) dihitung dengan menjumlahkan dari Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non Makanan (GKNM) (Aprilia & Sembiring, 2021).

Di Provinsi Sumatera Utara, kemiskinan menjadi isu krusial yang memerlukan perhatian khusus, dikarenakan masih adanya komposisi populasi dengan angka kemiskinan yang mencapai angka 7,19 % atau setara dengan 1,11 juta jiwa. Garis kemiskinan Sumatera Utara pada September 2024 tercatat sebesar Rp.648.336,-/kapita/bulan dengan komposisi Garis Kemiskinan Makanan sebesar Rp.495.730,- (76,46%) dan Garis Kemiskinan Bukan Makanan sebesar Rp.152.606,- atau sekitar 23,54% (BPS Provinsi Sumatera Utara, 2024). Sedangkan angka kemiskinan ekstrem berada pada angka 0,78%, lebih rendah dari rata-rata angka kemiskinan ekstrem nasional yang berada pada angka 0,83%.

Kemiskinan ekstrem secara global dan nasional mengalami tren penurunan, namun kendati demikian menurut data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE), estimasi penduduk yang mengalami kemiskinan ekstrem pada tahun 2024 di Provinsi Sumatera Utara

mengalami peningkatan 0,02% (Noor Achmad et al., 2024). Berdasarkan fakta bahwa kemiskinan ekstrem adalah titik terendah dalam spektrum kemiskinan maka diperlukan upaya penanganan yang serius dan berkelanjutan agar masalah ini cepat teratasi dan Sumatera Utara segera terbebas dari kemiskinan ekstrem serta dapat bertransisi ke tingkat kemiskinan yang lebih rendah.

Dengan mempertimbangkan adanya kenaikan kemiskinan ekstrem ini, maka penting untuk menganalisis apakah Provinsi Sumatera Utara berpotensi untuk bertransisi ke tingkat kemiskinan yang lebih buruk lagi dan bagaimana tingkat kemiskinan ini akan diproyeksikan untuk tahun-tahun berikutnya. Penelitian semacam ini belum banyak dilakukan, terlebih apabila ditinjau berdasarkan dimensi kedalaman dan keparahan kemiskinan suatu wilayah untuk bisa bertransisi melalui analisis dengan pendekatan model matematis.

Dalam kemiskinan absolut, tingkat kemiskinan dapat diukur dengan menggunakan Indeks Kedalaman Kemiskinan (*Poverty Gap Indeks - P_1*) dan Indeks Keparahannya Kemiskinan (*Poverty Severity Index - P_2*) yang dikembangkan oleh Foster-Greer-Thorbecke pada tahun 1984 (World Bank, 2005). Indeks Kedalaman Kemiskinan menjelaskan seberapa jauh jarak rata-rata pengeluaran penduduk dibandingkan garis kemiskinan sedangkan Indeks Keparahannya Kemiskinan menjelaskan seberapa besar ketimpangan atau ketidakmerataan sebaran pengeluaran diantara penduduk miskin. Diartikan (Tan & Syafri, 2024) sebagaimana dikutip dari (World Bank, 2005), semakin tinggi nilai dari indeks-indeks ini, maka semakin sulit pula untuk suatu wilayah bisa keluar dari kemiskinan.

Berdasarkan data September 2024, Indeks Kedalaman Kemiskinan Provinsi Sumatera Utara mencapai angka 1,23%, lebih rendah dibandingkan Indeks Kedalaman Kemiskinan nasional yang berada pada angka 1,364%, namun ironinya terdapat 12 kabupaten/kota yang berada di bawah rata-rata P_1 provinsi, dengan kemiskinan terdalam terjadi di Nias Utara yang bernilai 10 kali lipat lebih besar dari minimum kedalaman kemiskinan kabupaten/kota di Sumatera Utara. Sedangkan Indeks Keparahannya Kemiskinan Sumatera Utara berada pada angka 0,30%, ini juga lebih rendah dari Indeks Keparahannya Kemiskinan nasional yang berada pada angka

0,322% namun terdapat pula 7 kabupaten/kota yang berada dibawah rata-rata provinsi, dengan nilai maksimum keparahan kemiskinan juga terjadi di Nias Utara, yang bernilai 16 kali lipat lebih besar dari nilai minimumnya (Badan Pusat Statistik, 2024b). Artinya dari data tersebut diketahui sebaran kemiskinan tidak merata di Provinsi Sumatera Utara, maka perlu dilakukan segmentasi pada kabupaten/kota yang ada di Sumatera Utara supaya dapat diketahui daerah mana yang paling rentan dan pola sebarannya untuk tiap-tiap tingkat kemiskinan. Karena data yang digunakan tidak berlabel maka pengklasteran yang sesuai dan paling efisien secara komputasi adalah *K-Means Clustering*.

K-Means Clustering merupakan jenis metode *clustering* tanpa adanya label (*Unsupervised learning*) yang melakukan pengelompokan data untuk mendapatkan karakteristik yang sama dengan sistem partisi berbasis jarak rata-rata data ke *centroid* (Huang et al., 2024). Secara persis penelitian dengan menggunakan dimensi kedalaman dan keparahan sebagai fitur penjelas kemiskinan di Provinsi Sumatera Utara belum dilakukan, maka *clustering* ini masih relevan untuk dilakukan. Hasil klasterisasi ini nantinya dapat dipergunakan untuk menggambarkan pola sebaran kemiskinan, serta dapat juga dipergunakan untuk memproyeksikan transisi tingkat kemiskinan dengan menganalisisnya lebih lanjut menggunakan Rantai Markov (*Markov Chain*).

Rantai Markov (*Markov Chain*) merupakan proses stokastik dimana keadaan masa depan dapat diprediksi dengan menggunakan data keadaan masa sekarang (Lestari & Jasuni, 2023). Transisi tingkat kemiskinan ini mengikuti proses stokastik, artinya nilai peluangnya berubah-ubah seiring waktu dalam suatu ketidakpastian, maka perubahan tingkat kemiskinan ini hanya dapat dijelaskan berdasarkan struktur peluang. Hal tersebutlah yang tepatnya dapat dilakukan oleh Rantai Markov dan dikarenakan transisi ini bergerak dalam langkah waktu diskrit maka jenis Rantai Markov yang digunakan adalah DTMC (*Discrete Time Markov Chain*).

Penelitian ini akan diuraikan kedalam dua bagian besar yakni *clustering* atau segmentasi tingkat kemiskinan dan analisis perubahan probabilitas dalam sistem yang dinamis.

KAJIAN TEORI

Kemiskinan Absolut

Definisi yang paling umum untuk menggambarkan kemiskinan adalah kurangnya sumber daya untuk memenuhi kebutuhan yang diukur dengan membandingkan pendapatan atau konsumsi individu dengan garis kemiskinan tertentu (Haughton & Khandker, 2009).

Berdasarkan cara pengukurannya kemiskinan dibagi menjadi dua jenis, salah satunya adalah kemiskinan absolut. Kemiskinan absolut adalah kondisi dimana seseorang atau sekelompok orang tidak dapat memenuhi kebutuhan dasar seperti makanan, pakaian, kesehatan, perumahan, dan lainnya (Permatasari, 2024). Dalam kemiskinan absolut, penduduk miskin didefinisikan sebagai penduduk dengan rata-rata pengeluaran per kapita per bulan (y_i) di bawah garis kemiskinan (GK), artinya saat $y_i < GK$ maka orang tersebut terbilang sebagai penduduk miskin. Kemiskinan absolut berlaku secara global sehingga dapat digunakan untuk membandingkan negara atau wilayah regional spesifik (World Bank, 2005).

Garis Kemiskinan (GK) dan Kemiskinan Ekstrem

Garis kemiskinan (GK) atau batas kemiskinan adalah tingkat minimum pendapatan yang dianggap perlu dipenuhi untuk memperoleh standar hidup yang mencukupi di suatu negara atau wilayah (Aprilia & Sembiring, 2021) dengan nilai ini dihitung dengan rumus sebagai berikut:

$$GK = GKM + GKNM$$

Dimana:

GK : Garis Kemiskinan

GKM : Garis Kemiskinan Makanan

$GKNM$: Garis Kemiskinan Non Makanan

Saat seseorang atau sekelompok orang tidak dapat memenuhi garis kemiskinan terkecil (Garis Kemiskinan Ekstrem) maka mereka disebut mengalami kemiskinan ekstrem, artinya mereka hidup dalam kondisi sangat miskin dan tidak mempunyai kemampuan dalam memenuhi kebutuhan paling dasar seperti makanan, air bersih dan tempat tinggal yang layak.

Ukuran-ukuran Kemiskinan

Terdapat beberapa alternatif dalam mengukur kemiskinan, namun yang paling umum digunakan,

yang juga digunakan secara luas di Indonesia adalah *headcount index* (P_0), *poverty gap index* (P_1), *poverty severity index* (P_2) (Badan Pusat Statistik, 2024a).

1) The Headcount Index (P_0)

Ukuran ini menjelaskan persentase dari populasi yang hidup di bawah garis kemiskinan yang secara matematis dirumuskan sebagai :

$$P_0 = \frac{1}{N} \sum_{i=1}^N I(y_i < z)$$

Dimana :

P_0 : Persentase penduduk miskin

N : Jumlah penduduk

y_i : Rata-rata pengeluaran per kapita

$I(.)$: Fungsi indikator boolean, apabila ekspresi . benar maka dihitung 1

z : Garis Kemiskinan (GK)

Indeks ini tidak memperhitungkan intensitas kemiskinan atau menjelaskan seberapa miskin penduduk miskin tersebut namun hanya menghitung proporsi penduduk yang miskin dalam suatu wilayah (World Bank, 2005), dengan 1 dihitung sebagai miskin dan 0 dihitung sebagai tidak miskin.

2) The Poverty Gap Index (P_1)

Indeks kedalaman kemiskinan atau *poverty gap index* adalah ukuran kemiskinan yang menjelaskan seberapa jauh jarak rata-rata pengeluaran ataupun pendapatan penduduk dari garis kemiskinan. yang dituliskan sebagai :

$$P_1 = \frac{1}{N} \sum_{i=1}^N \frac{G_i}{z}$$

dengan kesenjangan kemiskinan (G_i) selalu non negatif didefinisikan sebagai:

$$G_i = (z - y_i) \times I(y_i < z)$$

Semakin tinggi indeks kedalaman kemiskinan ini maka semakin kecil persentase dari garis kemiskinan yang dapat dipenuhi. Nilai indeks ini berada di rentang 0 ke 1, dengan nilai 0 tercapai saat tidak ada penduduk yang berada dibawah dan garis kemiskinan dan nilai 1 tercapai saat penduduk dibawah garis kemiskinan tidak mempunyai penghasilan sama sekali.

3) The Poverty Severity Index (P_2)

Indeks Keparahan Kemiskinan atau *poverty severity index* menjelaskan seberapa besar ketimpangan atau ketidakmerataan pendapatan diantara penduduk miskin. Sama halnya dengan indeks kedalaman kemiskinan, nilai indeks ini berada di rentang 0 ke 1, dimana semakin tinggi nilai indeks mendekati 1 maka semakin besar pula ketidakmerataan pendapatan. Indeks ini dirumuskan sebagai berikut :

$$P_2 = \frac{1}{N} \sum_{i=1}^N \left(\frac{G_i}{z} \right)^2$$

Semakin tinggi Indeks Kedalaman Kemiskinan maka Indeks Keparahan kemiskinan ini juga akan semakin tinggi (Habib & Wahyudi, 2022).

Transisi Kemiskinan

Transisi kemiskinan diukur sebagai perubahan status atau tingkat kemiskinan dari tahun sebelumnya ($Y - 1$) ke tahun berikutnya ($Y0$) (Youn et al., 2020). Dalam kemiskinan absolut secara sederhana, transisi kemiskinan ini menjelaskan perpindahan pendapatan penduduk yang sebelumnya berada dibawah garis kemiskinan menjadi diatas garis kemiskinan (Stevens, 2012).

K-Means Clustering

K-Means Clustering merupakan jenis metode *clustering* yang paling populer dalam *Machine Learning* dengan tanpa adanya label (*Unsupervised learning*) yang melakukan pengelompokan data untuk mendapatkan karakteristik yang sama dengan sistem partisi berbasis jarak rata-rata data ke *centroid* (Huang et al., 2024). Dalam klasterisasi ini data akan dikelompokkan ke dalam $k < n$ cluster berdasarkan pola kesamaannya, dimana n adalah banyaknya data. Umumnya untuk jarak rata-rata data ke *centeroid* ini dihitung menggunakan jarak *euclidian*, yang ditunjukkan oleh:

$$D_e = \sqrt{(x_i - p_i)^2 + (y_i - q_i)^2 + \dots}$$

Dimana:

D_e : Jarak *euclidian*

i : banyaknya data/objek

x, y : Koordinat data/objek

p, q : Koordinat *centroid*.

Metode ini mencari *centroid* terbaik dengan memperbaharui letak *centroid* pada setiap iterasi berdasarkan jarak rata-rata, dan iterasi akan berhenti saat tidak terjadi lagi perpindahan *centroid* atau hanya terjadi perpindahan yang minimal dan cenderung konvergen ke suatu titik.

Langkah-langkah untuk melakukan *K-Means Clustering* adalah sebagai berikut:

1. Tentukan jumlah kluster (k) yang diinginkan
2. Inisiasi *centroid* awal secara acak atau pilih random dari objek
3. Untuk semua data, hitung jarak data ke *centroid* (dengan *Euclidean Distance*) dan alokasikan data ke *centroid* terdekatnya
4. Hitung rata-rata fitur data *cluster* yang sama dan perbaharui posisi *centroid*
5. Lakukan iterasi menghitung jarak data pada langkah 3 dan 4
6. Saat tidak ada perpindahan posisi *centroid* atau hanya terjadi perpindahan yang tidak signifikan maka proses iterasi berhenti.

Elbow Method

Dalam *K-Means Clustering*, sebelum melakukan klasterisasi harus ditentukan berapa nilai k . Nilai k dapat dipilih sesuai dengan jumlah *cluster* yang diinginkan, namun tidak ada jaminan hasil klasterisasi akan selalu optimal. Jadi digunakanlah *Elbow Method* untuk memastikan nilai k yang terpilih adalah jumlah kluster optimal. Metrik dalam menentukan jumlah kluster ini dilihat dari nilai *WCSS* yang apabila nilai k semakin tinggi maka nilai *WCSS* akan semakin rendah. Umumnya pemilihan k ini dilihat melalui grafik *WCSS (Within Cluster Sum of Squares)* dimana nilai k optimal ditunjukkan oleh titik siku dimana penurunan *WSS* mulai melambat secara signifikan (Noviandy et al., 2024). Dengan *WSS* dirumuskan dengan :

$$WCSS = \sum_{k=1}^k \sum_{x_i \in C_k} \|x_i - \mu_k\|^2$$

Dimana:

WCSS : Inersia atau *Within Cluster Sum of Squares*

k : Jumlah *cluster*

x_i : Data dalam *cluster*

μ_k : *Centroid cluster k*

C_k : *Cluster k*

Klaster yang baik dan berkualitas dalam *K-Means Clustering* setidaknya mempunyai dua properti, yakni:

1. Setiap titik data dalam satu *cluster* memiliki kesamaan yang tinggi
2. Data antar *cluster* memiliki perbedaan yang tinggi

Kedua sifat yang menjelaskan kinerja clustering tersebut dapat dievaluasi melalui rumus sederhana yakni:

$$Total_SS = WCSS + Between_SS$$

Dimana:

Total_SS (Total Sum of Squares) : Total Variansi dalam data atau jumlah kuadrat jarak setiap unit data ke *centroid* keseluruhan data

WCSS : Ukuran kerapatan data/homogenitas dalam satu *cluster* atau jumlah kuadrat jarak seluruh data dalam satu *cluster* ke *centroidnya*

Between_SS : Ukuran seberapa baik *cluster-cluster* terpisah satu sama lain

Semakin rendah nilai *WCSS* (mendekati 0) maka homogenitas data dalam *cluster* yang sama akan semakin tinggi, namun semakin rendah *Between_SS* maka semakin buruk *cluster-cluster* terpisah satu sama lain artinya kualitas hasil *clustering* semakin menurun (Kaufman & Rousseeuw, 1989)

Deteksi Outlier Z-Score

Outlier atau data pencilan adalah data yang terletak jauh dari mayoritas data lainnya. Umumnya pada tahap *Pre-processing* data, dilakukan pendeteksian *outlier* untuk menghindari kesalahan analisis dan meningkatkan kualitas model (Agustianto et al., 2020). *Z-Score* setiap unit data dapat dihitung dengan menggunakan rumus :

$$Z = \frac{x - \mu}{\sigma}$$

Dengan :

x : Nilai data setiap unit dari fitur

μ : Rata-rata fitur

σ : Standar Deviasi fitur

Rentang kenormalan atau interval nilai *Z-Score* agar data tidak terdeteksi sebagai *outlier* adalah: $-3 \leq Z \leq 3$

Maka data akan terdeteksi *outlier* apabila $|Z| > 3$. Ini didasarkan pada pada sifat distribusi normal standar (*bell curve*) dimana 99.7% data berada

dalam rentang $z = \pm 3$. Dengan mengikuti aturan 3 sigma artinya hanya 0.3% data (3 dari 1000 data) yang diharapkan berada di luar $z = \pm 3$ pada distribusi normal. Nilai di luar ini dianggap jarang/ekstrem (Shiffler, 1988).

K-Means Sharp (K-Means#)

K-Means klasik memilih dan memperbaharui *centroid* berdasarkan rata-rata, oleh karena itu K-Means sangat sensitif terhadap *outlier*, hal tersebutlah yang menjadi salah satu kelemahan terbesar dari K-Means karena *outlier* dalam dataset dapat menarik *centroid* kluster menjauhi kelompok data yang sebenarnya, menghasilkan pengelompokan yang kurang akurat. Maka untuk mengatasi hal tersebut dirancanglah variasi dari K-Means yang lebih tahan terhadap *outlier* yang disebut K-Means Sharp (K-Means #). K-Means Sharp adalah modifikasi dari K-Means yang mengubah metode pembaruan *centroid* dengan mengisolasi *outliers* sehingga pemilihan inisial dan pembaharuan *centroid* terbebas dari intervensi *outliers*. Karena *outlier* dikecualikan dari perhitungan *centroid*, kluster yang dihasilkan oleh K-Means# cenderung lebih representatif dari kelompok data yang sebenarnya. *Centroid* tidak "terdistorsi" oleh keberadaan *outlier*, sehingga batas-batas kluster menjadi lebih jelas dan akurat (Olukanmi & Twala, 2021).

Rantai Markov (Markov Chain)

Rantai Markov adalah suatu proses stokastik yang menggambarkan transisi atau perpindahan keadaan (*State*) di mana probabilitas setiap kejadian hanya bergantung pada keadaan yang dicapai pada kejadian sebelumnya, artinya masa depan dari proses tersebut hanya bergantung pada kondisi saat ini dan tidak dipengaruhi oleh keadaan sebelumnya (Ross, 2010). Sedangkan *Discrete Time Markov Chain (DTMC)* adalah proses Markov dengan ruang keadaan merupakan himpunan terbatas dan dapat dihitung (*countable*), dengan himpunan waktu ($T = 0, 1, 2, \dots, t$). Secara umum suatu proses Markov memenuhi sifat Markovian (*Memorylessness*) yakni

$$\begin{aligned} P_r\{X_{t+1} = j | X_0 = i_0, \dots, X_{t-1} = i_{t-1}, X_t = i\} \\ = P_r\{X_{t+1} = j | X_t = i\} \end{aligned}$$

Untuk semua titik waktu t dan dengan keadaan $i_0, \dots, i_{t-1}, i, j$.

Peluang X_{n+1} pada keadaan j dan peluang X_n dalam keadaan i , disebut sebagai peluang transisi satu langkah (*one-step probability*) dan dilambangkan dengan $P_{ij}^{t,t+1}$ dengan

$$P_{ij}^{t,t+1} = P\{X_{t+1} = j | X_t = i\}$$

Matriks Peluang Transisi

Misalkan sistem terdefinisi kedalam n state, seluruh peluang transisi state (P_{ij}) disusun dalam sebuah matriks persegi $n \times n$ yang disebut Matriks Peluang Transisi (*Transition Probability Matrix*). Matriks peluang transisi adalah matriks stokastik yang mengatur perpindahan sistem dari satu keadaan ke keadaan lainnya umumnya dinotasikan sebagai P atau P^n (Ross, 1996) dan secara matematis ditulis sebagai:

$$P = P^n = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \dots \\ P_{10} & P_{11} & P_{12} & \dots \\ \vdots & \vdots & \vdots & \ddots \\ P_{i0} & P_{i1} & P_{i3} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Dimana semua P_{ij} harus bernilai non negatif dan memenuhi kondisi:

1. $0 \leq P_{ij} \leq 1$, dengan i dan j ; $n = 0, 1, 2, \dots, M$
2. $\sum_{j=0}^M P_{ij} = 1$ dengan i ; $n = 0, 1, 2, \dots, M$

Dengan nilai peluang transisi dihitung dengan rumus umum peluang yakni :

$$P_{ij} = \frac{n_{ij}(t)}{n_i(t)}$$

Dimana:

P_{ij} : Peluang transisi dari state i ke state j

$n_{ij}(t)$: Banyaknya objek yang bertransisi dari state i ke state j dalam periode t

$n_i(t)$: nilai state i dalam periode t

Persamaan Chapman-Kolmogorov

Definisikan n langkah probabilitas transisi P_{ij}^n sebagai probabilitas yang menjelaskan sebuah proses di state i akan berada di state j setelah sejumlah transisi

$$P_{ij}^n = P\{X_{t+k} = j | X_k = i\}, n \geq 0, i, j \geq 0$$

Persamaan *Chapman-Kolmogorov* menyediakan metode untuk menghitung n langkah probabilitas transisi ini yang dituliskan secara matematis sebagai:

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^n P_{kj}^m, \forall n, m \geq 0$$

$P_{ik}^n P_{kj}^m$ merepresentasikan probabilitas yang berawal dari i bertransisi ke $state j$ dalam $n + m$ langkah transisi melalui $state k$ dengan sebanyak n transisi. Oleh karena itu, menjumlahkan semua $states$ diantara k menghasilkan probabilitas yang menjelaskan proses di $state j$ setelah $n + m$ transisi (Ross, 2010). Secara formal dituliskan

$$\begin{aligned} P_{ij}^{n+m} &= P\{X_{n+m} = j | X_0 = i\} \\ &= \sum_{k=0}^{\infty} P\{X_{n+m} = j, X_n = k | X_0 = i\} \\ &= \sum_{k=0}^{\infty} P\{X_{n+m} = j | X_n = k, X_0 = i\} P\{X_n = k | X_0 = i\} \\ &= \sum_{k=0}^{\infty} P_{kj}^m P_{ik}^n \end{aligned}$$

Jika P^n menotasikan matriks probabilitas transisi n langkah, maka persamaan Chapman-Kolmogorov menunjukkan bahwa :

$$P^{(n+m)} = P^{(n)} \cdot P^{(m)}$$

Dimana dot menunjukkan perkalian matriks karenanya secara khusus

$$P^{(2)} = P^{(1+1)} = P \cdot P = P^2$$

Dan dengan induksi

$$P^{(n)} = P^{(n-1+1)} = P^{n-1} \cdot P = P^n$$

Artinya matriks probabilitas transisi n langkah dapat diperoleh dengan mengalikan matriks P sebanyak n kali.

Peluang Steady State

Proses Markov umumnya akan menuju suatu kondisi *Steady State* (keseimbangan) artinya setelah proses berjalan selama beberapa periode, probabilitas yang dihasilkan akan bernilai tetap dan tidak lagi berubah untuk transisi waktu kedepannya (Ross, 2010). Dinotasikan sebagai:

$$\pi_j = \lim_{n \rightarrow \infty} P_{ij}^{(n)}$$

Dimana

π_j : Peluang *steady state*

$P_{ij}^{(n)}$: Peluang transisi dari $state i$ ke j dalam n langkah

Dengan peluang *steady state* ini memenuhi sifat-sifat:

$$\begin{aligned} \pi_j &= \sum_{i=0}^{\infty} P_{ij}^{(n)} \pi_i, j = 0, 1, \dots, M \\ \sum_{j=0}^M \pi_j &= 1 \end{aligned}$$

Dengan mempertimbangkan probabilitas *steady state* yang menyusun matriks probabilitas transisi, maka pada kondisi *steady state*, matriks ini dapat disajikan sebagai vektor v yang memenuhi :

$$\begin{aligned} vP &= v \\ v(P - I) &= 0 \end{aligned}$$

Dengan I adalah matriks Identitas.

Untuk menjamin kekonvergenan dan keunikan nilai peluang jangka panjang bila berangkat dari sembarang *state* awal dilakukan uji *state ergodic*. *Ergodic Markov Chain* memenuhi sifat *irreducible*, *aperiodic* dan *positive recurrent*.

1. Irreducible

Suatu rantai Markov disebut *irreducible* (tak tereduksi) jika setiap keadaan saling berkomunikasi dengan setiap keadaan lain atau dengan kata lain hanya memiliki satu kelas komunikasi (Nasib et al., 2024).

2. Aperiodic

State i dapat dikatakan *aperiodic* atau tidak memiliki pola periodik dalam transisi, jika dan hanya jika $d(i) = 1$, dimana $d(i) = \gcd\{n \geq 1, P_{ii}^n > 0\}$

3. Positive recurrent

Suatu *state* disebut berulang positif (*Positive Recurrent*) jika *state* tersebut *recurrent* serta berlaku: jika proses dimulai dari *state i* maka nilai harapan dari waktu sampai proses tersebut kembali ke *state i* ($\mu_i = \frac{1}{\pi_i}$) adalah bilangan hingga (*finite*) (Riyono et al., 2022).

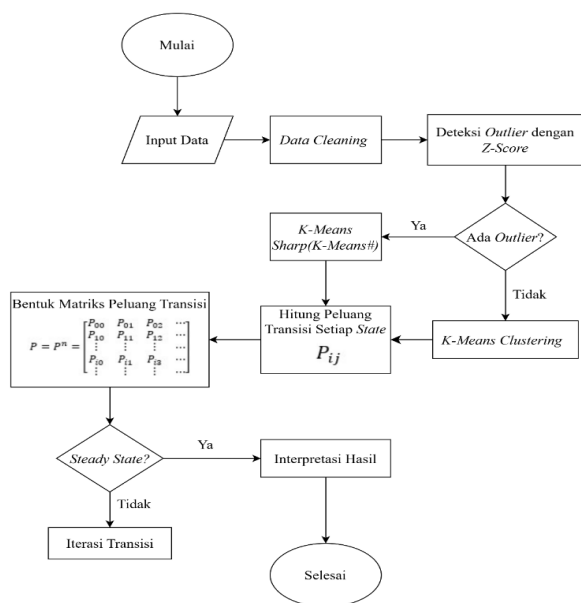
METODE

Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari *website* Badan Pusat Statistik Sumatera Utara yang mencakup data Indeks Kedalaman Kemiskinan (*Poverty Gap Indeks - P₁*) dan Indeks Keparahan Kemiskinan (*Poverty Severity Index - P₂*) kabupaten/kota di Provinsi Sumatera Utara tahun 2023-2024.

Analisis Data

Analisis data dalam penelitian ini terdiri dari beberapa tahapan yang secara ringkas disajikan dalam diagram alir berikut :



Gambar 1. Diagram Alir Analisis Data

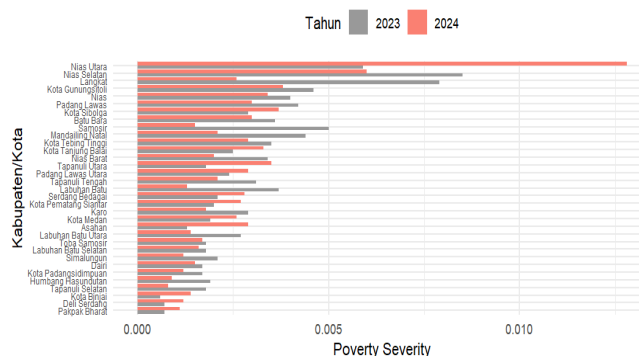
HASIL DAN PEMBAHASAN

Deskripsi Umum Data

Berdasarkan data September 2024 yang diukur menggunakan pendekatan kemiskinan absolut atau kebutuhan dasar minimum di Provinsi Sumatera Utara masih terdapat komposisi populasi dengan angka kemiskinan sebesar 7,19 % atau setara dengan 1,11 juta jiwa berada dibawah garis kemiskinan. Garis kemiskinan Sumatera Utara tercatat sebesar Rp.648.336,-/kapita/bulan dengan komposisi Garis Kemiskinan Makanan sebesar Rp.495.730,- (76,46%) dan Garis Kemiskinan Bukan Makanan sebesar Rp.152.606,- atau sekitar 23,54% (BPS Provinsi Sumatera Utara, 2024). Sedangkan angka kemiskinan ekstrem berada pada angka 0,78% , lebih rendah dari rata-rata angka kemiskinan ekstrem nasional yang berada pada angka 0,83%.

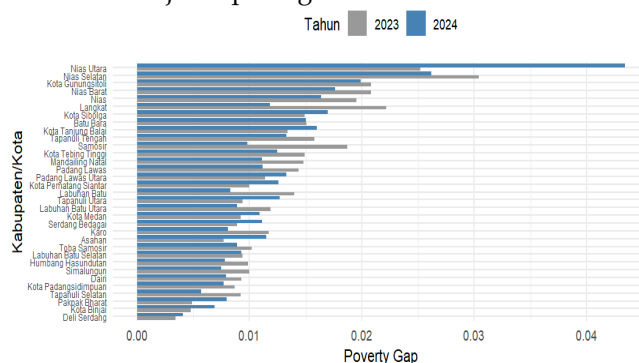
Pada periode September 2024, Indeks Kedalaman Kemiskinan (*Poverty Gap- P₁*) Provinsi Sumatera Utara mencapai angka 1,23%, lebih rendah dibandingkan Indeks Kedalaman Kemiskinan Nasional yang berada pada angka 1,364%, namun ironinya terdapat 12 kabupaten/kota yang berada di bawah rata-rata *P₁* provinsi, dengan kemiskinan terdalam terjadi di Kabupaten Nias Utara yang bernilai 10 kali lipat lebih besar dari minimum kedalaman kemiskinan kabupaten/kota di Sumatera Utara yang dicapai oleh Kabupaten Deli Serdang. Sebaran *Poverty Gap* atau Indeks

Kedalaman Kemiskinan di Provinsi Sumatera Utara tahun 2023 dan 2024 disajikan pada grafik berikut:

Gambar 2. Sebaran *Poverty Gap* Sumut 2023 dan 2024

Gambar 2 menunjukkan adanya perubahan sebaran kedalaman kemiskinan di Provinsi Sumatera Utara pada transisi tahun 2023-2024, terjadi peningkatan kedalaman kemiskinan pada 12 wilayah kabupaten/kota dan penurunan pada 21 kabupaten/kota lainnya. Peningkatan kedalaman kemiskinan paling signifikan terjadi di Kabupaten Nias Utara dengan persentase kenaikan sebesar 72,2%, sementara penurunan paling signifikan terjadi di Kabupaten Samosir dengan persentase penurunan 47,59%.

Sama halnya dengan *P₁*, Indeks Keparahan Kemiskinan (*Poverty Severity- P₂*) Sumatera Utara berada pada angka 0,30%, juga lebih rendah dari Indeks Keparahan Kemiskinan Nasional yang berada pada angka 0,322% namun terdapat pula 7 kabupaten/kota yang berada dibawah rata-rata provinsi, dengan nilai maksimum keparahan kemiskinan juga terjadi di Nias Utara, yang bernilai 16 kali lipat lebih besar dari nilai minimumnya yang dicapai oleh Kabupaten Tapanuli Selatan. Sebaran *Poverty Severity* atau Indeks Keparahan Kemiskinan di Provinsi Sumatera Utara tahun 2023 dan 2024 disajikan pada grafik berikut:

Gambar 3. Sebaran *Poverty Severity* Sumut 2023 dan 2024

Gambar 3 menunjukkan terjadinya perubahan sebaran keparahan kemiskinan di Provinsi Sumatera Utara pada transisi tahun 2023-2024, terjadi peningkatan keparahan kemiskinan pada 12 wilayah kabupaten/kota dan penurunan pada 21 kabupaten/kota lainnya.. Secara ringkas informasi umum data disajikan pada tabel statistik deskriptif berikut :

Tabel 1. Statistik Deskriptif

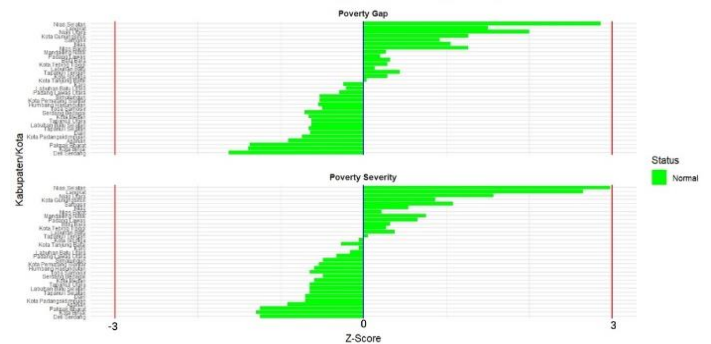
Ukuran Statistik	Poverty Gap		Poverty Severity	
	2023	2024	2023	2024
Mean	0,01317	0,01249	0,00300	0,00262
Median	0,0117	0,0111	0,0025	0,0021
Mode	0,0092	0,0111	0,0018	0,0029
Std Deviasi	0,006008	0,007133	0,001845	0,002131
Count	33	33	33	33

Berdasarkan Tabel 1 dapat diketahui, bahwa secara keseluruhan rata-rata *Poverty Gap* dan *Poverty Severity* di Provinsi Sumatera Utara sama-sama mengalami penurunan. Pada tahun 2023 nilai rata-rata P_1 adalah 0,0131 kemudian menurun sebesar 5,17% menjadi 0,0124 di tahun 2024. Turunnya nilai P_1 ini mengindikasikan adanya kecenderungan peningkatan rata-rata pengeluaran konsumsi penduduk miskin, atau dengan kata lain selisih pengeluaran penduduk miskin terhadap garis kemiskinan semakin berkurang. Nilai rata-rata P_2 Sumatera Utara pada tahun 2023 adalah 0,003 kemudian menurun sebesar 12,51% menjadi 0,0026, artinya ketimpangan pengeluaran konsumsi diantara penduduk miskin semakin berkurang, atau dengan kata lain penyebaran pengeluaran konsumsi semakin baik atau merata.

Clustering Data 2023

1) Deteksi *Outlier*

Sebelum melakukan klasterisasi umumnya selalu dilakukan eksplorasi untuk mendeteksi ada tidaknya *outlier* dalam *dataset*. Pada praktiknya terdapat beberapa metode dalam pendeteksian *outlier*, salah satunya adalah menggunakan *Z-Score*. Dengan menggunakan persamaan dihitung nilai *Z-Score* setiap unit data dari kedua fitur, yang disajikan melalui grafik berikut:

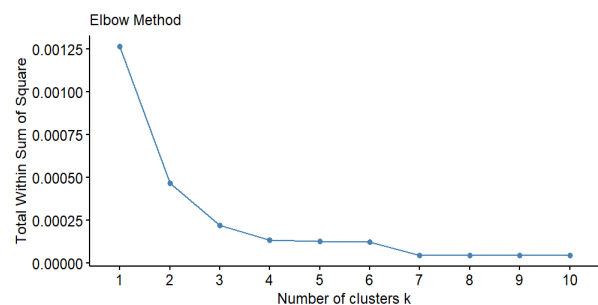


Gambar 4. Sebaran Nilai Z-Score Data 2023

Berdasarkan Gambar 4 dapat diketahui bahwa pada data tahun 2023 kedua fitur yakni P_1 dan P_2 bersih dari *outlier* dengan nilai *Z-Score* maksimum bernilai positif sebesar 2,86 untuk P_1 dan 2,97 untuk P_2 mendekati batas 3 yang ditetapkan sebagai batas kenormalan.

2) Jumlah *Cluster* Optimal dengan *Elbow Method*

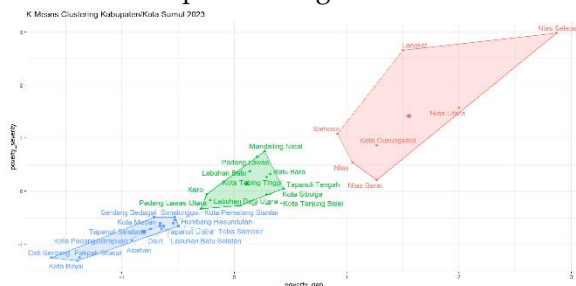
Metode *elbow* (sikut) memilih nilai k atau jumlah *cluster* yang optimal dengan interpretasi melalui grafik *WCSS* (*Within Cluster Sum of Squares*) dimana nilai k optimal ditunjukkan oleh titik siku dimana penurunan *WCSS* secara drastis dan mulai melambat secara signifikan setelahnya. (Noviandy et al., 2024). Penentuan nilai k dengan *elbow method* dapat dilihat melalui grafik berikut:

Gambar 5. Jumlah *Cluster* Optimal Data 2023 dengan *Elbow Method*

Berdasarkan Gambar 5 dapat diketahui bahwa nilai k optimal adalah 3 karena saat di titik $k = 3$, penurunan inersia masih terjadi secara signifikan dan setelahnya penurunan nilai inersia ini mulai melambat atau mulai melandai, juga merupakan titik yang paling membentuk *elbow* atau sikut. Maka data 2023 ini akan diklasterkan dengan *K-Means Clustering* ke dalam 3 *cluster* yang akan diinisiasi dengan 3 *centroid* awal juga.

3) K-Means Clustering

Berdasarkan kaidah *Elbow Method* diatas, jumlah *cluster* optimal adalah 3 maka algoritma *K-Means Clustering* diawali dengan menginisiasi 3 *centroid* awal secara acak. Kemudian dengan rumus jarak *euclid* akan dihitung jarak *Euclid* setiap data ke *centroid* dan mengelompokkannya ke *centroid* terdekatnya, maka akan terbentuk *cluster-cluster* kemiskinan pada satu iterasi. Setelahnya akan dihitung rata-rata fitur setiap data pada *cluster* yang sama, kemudian menjadikannya sebagai *centroid* pada iterasi berikutnya. Proses ini berlanjut sampai tidak terjadi perpindahan posisi *centroid* yang signifikan atau anggota *cluster* tidak lagi berpindah. Visualisasi kelompok *cluster* kemiskinan saat iterasi berhenti ditampilkan sebagai berikut:



Gambar 6. Visualisasi *Clustering* Kemiskinan Sumatera Utara pada Tahun 2023

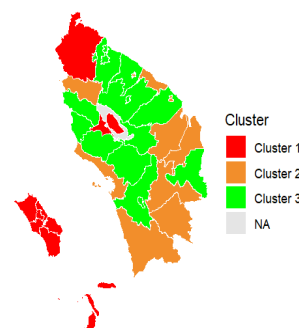
Berdasarkan hasil klasterisasi kemiskinan Sumatera Utara 2023 menggunakan *K-Means Clustering*, ditemukan bahwa data akan terkelompok kedalam 3 *cluster* optimal yang masing-masing diisi oleh 7 kabupaten/kota pada *cluster* 1, 11 kabupaten/kota pada *cluster* 2 dan 15 kabupaten/kota pada *cluster* 3. *Cluster* 1 adalah *cluster* dengan nilai *centroid* tertinggi artinya *cluster* ini adalah kabupaten/kota dengan rata-rata kekurangan pendapatan terhadap garis kemiskinan terbesar di Sumatera Utara dan juga merupakan kabupaten/kota dengan sebaran pendapatan paling tidak merata. *Cluster* 2 adalah *cluster* dengan tingkat kemiskinan menengah dengan rata-rata *poverty gap* sebesar 0,0138 dan rata-rata *poverty severity* sebesar 0,0032. Sedangkan *cluster* 3 adalah *cluster* tingkat kemiskinan dengan nilai *centroid* terendah, artinya kabupaten/kota yang berada dalam kelompok ini memiliki rata-rata kekurangan pendapatan terhadap garis kemiskinan yang terendah serta memiliki sebaran pengeluaran perkapita yang paling merata di Sumatera Utara. Setelah menjalankan algoritma sebanyak 100 kali melalui program R diperoleh

initial centroid dan nilai *Within Cluster Sum of Squares* (WCSS) tiap-tiap *cluster* seperti berikut yang disajikan pada Tabel 2.

Tabel 2. *Initial Centroids Cluster* Tahun 2023

Cluster	Centroid		WCSS
	Poverty Gap	Poverty Severity	
1	0,022514286	0,005614286	0.000121557
2	0,013845455	0,003263636	0.000027892
3	0,008333333	0,001593333	0.000069402

Hasil klasterisasi menemukan bahwa *Total Sum of Squares* optimal adalah sebesar 0,0002188525. Dengan *Between Sum of Squares* atau seberapa baik *cluster-cluster* terpisah satu sama lain sebesar 82,7%. Berdasarkan hasil klasterisasi Sumatera Utara tahun 2023 diatas, dilakukan pemetaan yang ditunjukkan pada gambar berikut, dengan NA merepresentasikan luasan Danau Toba.



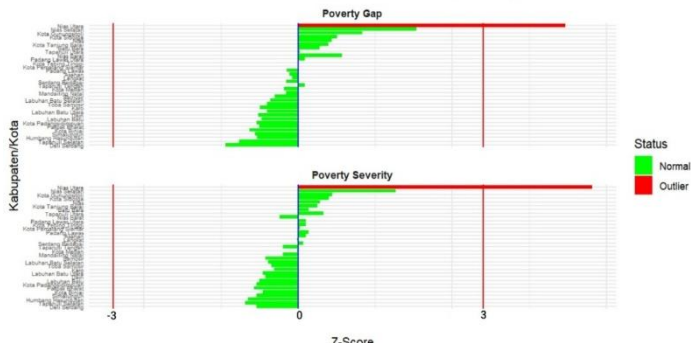
Gambar 7. Pemetaan Hasil *K-Means Clustering* Sumut 2023

Clustering Data 2024

1) Deteksi *Outlier*

K-Means Clustering adalah metode analisis *cluster* yang pada penerapannya memperbaharui *centroid* berdasarkan rata-rata, yang mana rata-rata bukanlah ukuran statistik yang *robust* atau tahan *outlier* seperti median, sehingga *K-Means* ini sangat sensitif dengan *outlier*. Maka sebelum melakukan klasterisasi umumnya selalu dilakukan eksplorasi untuk mendeteksi ada tidaknya *outlier* dalam dataset. Pada praktiknya terdapat beberapa metode dalam pendeteksian *outlier*, salah satunya adalah menggunakan *Z-Score*.

Pada data 2024 juga dilakukan pendeteksian *outlier* dengan metode yang sama yakni dengan *Z-Score*. Sebaran *Z-Score* setiap unit data dari kedua fitur, disajikan melalui grafik berikut:

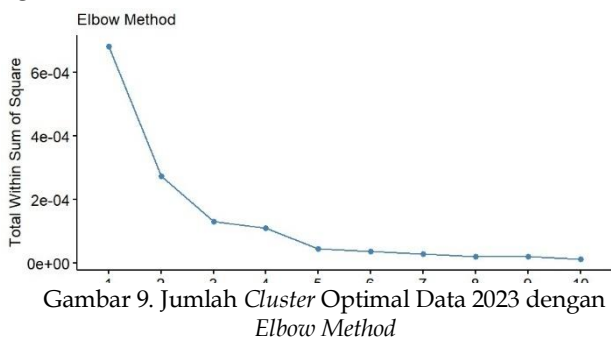


Gambar 8. Sebaran Nilai *Z-Score* Data 2024

Berdasarkan Gambar 8. menunjukkan adanya data abnormal atau *outlier* dalam data tahun 2024 pada kedua fitur. Daerah yang terdeteksi sebagai *outlier* dan melewati batas kenormalan adalah Kabupaten Nias Utara dengan nilai *Z-Score* mencapai 4,33 untuk P_1 dan 4,77 untuk P_2 . Adanya *outlier* ini mengakibatkan *dataset* tidak cukup baik untuk diklasterkan secara langsung menggunakan *K-Means* klasik karena nantinya *outlier* ini akan menarik *centroid* kluster menjauhi kelompok data yang sebenarnya, menghasilkan pengelompokan yang kurang akurat. Maka selanjutnya untuk data 2024 akan diklasterkan menggunakan variasi dari *K-Means Clustering* yang mengisolasi *outlier* saat memilih dan memperbaharui *centroid* yakni *K-Means Sharp (K-Means#)* (Olukanmi & Twala, 2021).

2) Penentuan Jumlah *Cluster* Optimal

Dalam *Elbow Method*, nilai k dipilih untuk meminimalkan *Within Cluster Sum Of Squares* (WCSS). Semakin kecil nilai WCSS maka variasi data dalam cluster akan semakin kecil atau data-data dalam suatu *cluster* akan memiliki kesamaan yang tinggi. Pemetaan nilai inersia atau WCSS terhadap berbagai nilai k pada data 2024 dapat dilihat melalui grafik berikut:

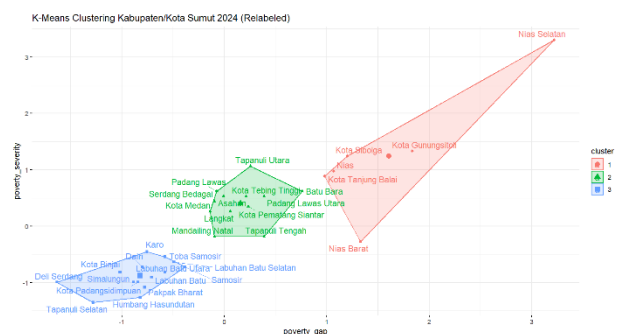


Gambar 9 menunjukkan bahwa jumlah *cluster* optimal pada data 2024 sama dengan jumlah *cluster* optimal pada data 2023 yakni 3 *cluster*. Pada titik

$k = 3$, penurunan inersia masih terjadi secara signifikan dan setelahnya penurunan nilai inersia ini mulai melambat atau mulai melandai, serta merupakan titik yang membentuk *elbow* atau siku. Kesamaan jumlah *cluster* ini nantinya akan menyusun matriks peluang transisi pada waktu transisi 2023-2024 yang lengkap dan persegi, artinya seluruh state yang ada akan saling berkomunikasi.

3) *K-Means Sharp Clustering*

Secara garis besar algoritma *K-Means Sharp* (*K-Means#*) hampir sama dengan algoritma *K-Means* klasik, perbedaan utamanya hanya terletak pada perlakuan terhadap data *outlier*. *K-Means* klasik mengasumsikan semua titik data sebagai data bukan pencilan, sehingga apabila terdapat data *outlier* maka data tersebut akan berkontribusi pada *Total Sum of Squares* yang menyebabkan hasil klasterisasi tidak akurat. *Outlier* akan otomatis menarik *centroid* kluster menjauhi kelompok data yang sebenarnya, maka untuk mengatasi hal tersebut dalam algoritma *K-Means Sharp* (*K-Means#*), dilakukan pengisolasian terhadap *outlier* saat pembaharuan *centroid*. Namun bukan berarti data *outlier* tersebut akan serta merta dihapus atau bahkan tidak akan mendapat label *cluster* yang artinya data tidak memiliki kemiripan dengan *cluster* manapun, tetapi data tersebut akan dialokasikan kedalam *cluster* dengan jarak Euclid terkecil dengan *centroid cluster* yang ditemukan tanpa adanya pengaruh *outlier*. Dikarenakan pada proses pendeteksian *outlier* dengan *Z-Score*, data *outlier* telah teridentifikasi dengan jelas maka pengaplikasian algoritma *K-Means Sharp* dapat dilakukan. Pada proses komputasi data Kabupaten Nias Utara yang merupakan *outlier* pada dataset 2024 tidak akan diikutsertakan yang mengakibatkan visualisasi klasterisasi seperti berikut :



Gambar 10. Visualisasi *Clustering* Kemiskinan Sumatera Utara pada Tahun 2024

Hasil klasterisasi menemukan bahwa *Total Sum of Squares* optimal pada data 2024 tanpa adanya pengaruh *outlier* adalah sebesar 0,0006821759. Dengan *Between Sum of Squares* atau seberapa baik *cluster-cluster* terpisah satu sama lain sebesar 81%. Melalui perulangan algoritma sebanyak 100 kali pada program R diperoleh *initial centroid* dan nilai *Within Cluster Sum of Squares (WCSS)* tiap-tiap *cluster* tanpa adanya pengaruh *outlier* seperti tabel berikut:

Tabel 3. *Initial Centroids Cluster Tahun 2024*

Cluster	Centroid		WCSS
	Poverty Gap	Poverty Severity	
1	0,01885	0,0037	0,000082675
2	0,01225	0,0027	0,000018139
3	0,00777	0,00132	0,000029132

Centroid yang ditampilkan diatas juga merupakan *centroid* pada saat algoritma *k-means* berhenti, karena kekonvergenan *centroid* terjadi pada iterasi kedua. Maka langkah selanjutnya dalam algoritma *K-Means Sharp* adalah untuk menugaskan *outlier* mencari *centroid* terdekatnya, dengan menggunakan rumus jarak *Euclid* diperoleh jarak data Kabupaten Nias Utara terhadap ketiga *centroid* sebagai berikut:

Tabel 4. Jarak Euclid *Outlier* terhadap Ketiga *Centroid*

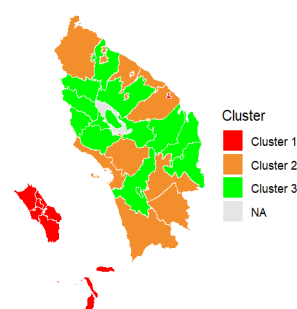
Centroid Cluster	Euclidian Distances
1	0.0261823
2	0.0327285
3	0.037423

Berdasarkan Tabel 4 ditemukan bahwa jarak Euclid terkecil data Nias Utara adalah dengan *cluster* 1 yakni sebesar 0,0261823, maka Nias Utara akan berada dalam *cluster* 1.

Berdasarkan hasil klasterisasi kemiskinan Sumatera Utara 2024 menggunakan *K-Means Sharp(K-Means#)*, ditemukan bahwa data juga terkelompok kedalam 3 *cluster* optimal yang masing-masing diisi oleh 7 kabupaten/kota pada *cluster* 1, 12 kabupaten/kota pada *cluster* 2 dan 14 kabupaten/kota pada *cluster* 3. *Cluster* 1 adalah *cluster* dengan nilai *centroid* tertinggi dengan rata-rata *poverty gap* sebesar 0,022357 yang berada dalam interval 0,016-0,0434 dan rata-rata *poverty severity*

sebesar 0,005 yang berada dalam interval 0,002-0,0128. *Cluster* 2 adalah *cluster* menengah dengan rata-rata *poverty gap* sebesar 0,01225 yang berada dalam interval 0,0109-0,0150 dan rata-rata *poverty severity* sebesar 0,002758 yang berada dalam interval 0,0021-0,0035. Sedangkan *cluster* 3 adalah *cluster* dengan nilai *centroid* paling rendah dengan rata-rata *poverty gap* sebesar 0,00777 yang berada dalam interval 0,0041-0,0098 dan rata-rata *poverty severity* sebesar 0,001328 yang berada dalam interval 0,00018-0,0018.

Berdasarkan hasil klasterisasi Sumatera Utara tahun 2024 diatas, diperoleh pemetaan sebagai berikut:

Gambar 11. Pemetaan Hasil *K-Means Clustering* Sumut 2024

Analisis Markov Chain

Dalam Rantai Markov, kondisi atau nilai kejadian yang dimiliki suatu sistem pada suatu waktu diwakili oleh *state* (keadaan). *State* ini merupakan himpunan bagian dari ruang kejadian yang dilambangkan dengan S . $S = \{S_n, n \geq 0\}$ dimana n merupakan banyaknya *state* dalam sistem. Dalam penelitian ini, *state* diwakili oleh *cluster-cluster* yang tercipta melalui proses *K-Means Clustering* dengan didefinisikan sebagai berikut:

State 1 : *Cluster* 1 atau *cluster* kemiskinan tinggi (ekstrem)

State 2 : *Cluster* 2 atau *cluster* kemiskinan menengah

State 3 : *Cluster* 3 atau *cluster* kemiskinan rendah.

1) Peluang Transisi dan Matriks Peluang Transisi

Peluang transisi adalah peluang proses berpindah dari satu *state* ke *state* lain dalam satu langkah waktu. Jika sebuah Rantai Markov mempunyai k kemungkinan *state* yang ditandai dengan $1, 2, \dots, k$, maka jika peluang *state* berada dalam keadaan j setelah mengalami kejadian i pada pengamatan sebelumnya dilambangkan dengan P_{ij} yang disebut

probabilitas transisi, dengan nilai P_{ij} dapat dihitung menggunakan persamaan :

$$P_{ij} = \frac{n_{ij}(t)}{n_i(t)}$$

Maka untuk memperoleh matriks peluang transisi, maka terlebih dahulu dihitung keseluruhan probabilitas transisinya yang ditunjukkan seperti berikut:

$$\begin{aligned} P_{11} &= \frac{n_{11}(t)}{n_1(t)} = \frac{5}{7} = 0,714 & P_{21} &= \frac{n_{21}(t)}{n_2(t)} = \frac{2}{11} = 0,182 & P_{31} &= \frac{n_{31}(t)}{n_3(t)} = \frac{0}{15} = 0 \\ P_{12} &= \frac{n_{12}(t)}{n_1(t)} = \frac{1}{7} = 0,143 & P_{22} &= \frac{n_{22}(t)}{n_2(t)} = \frac{6}{11} = 0,545 & P_{32} &= \frac{n_{32}(t)}{n_3(t)} = \frac{5}{15} = 0,333 \\ P_{13} &= \frac{n_{13}(t)}{n_1(t)} = \frac{1}{7} = 0,143 & P_{23} &= \frac{n_{23}(t)}{n_2(t)} = \frac{3}{11} = 0,273 & P_{33} &= \frac{n_{33}(t)}{n_3(t)} = \frac{10}{15} = 0,667 \end{aligned}$$

Secara ringkas probabilitas transisi seluruh *state* diringkas dalam tabel berikut:

Tabel 4.5. Peluang Transisi

Tingkat Kemiskinan	Ekstrem (1)	Menengah (2)	Rendah (3)	Total
Ekstrem (1)	0,714	0,143	0,143	1
Menengah (2)	0,182	0,545	0,273	1
Rendah (3)	0	0,333	0,667	1

Berdasarkan perhitungan peluang transisi diatas, dapat dibentuk matriks peluang transisi sebagai berikut:

$$P = \begin{bmatrix} 0,714 & 0,143 & 0,143 \\ 0,182 & 0,545 & 0,273 \\ 0 & 0,333 & 0,667 \end{bmatrix}$$

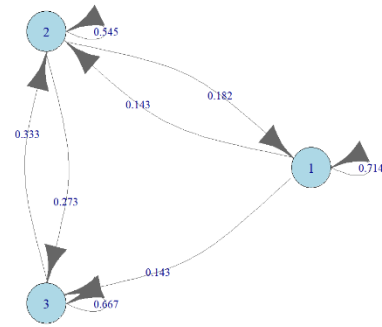
2) Peluang Steady State

Sebelum menganalisis distribusi *Steady state* perlu ditunjukkan bahwa Rantai Markov merupakan *Ergodic* untuk menjamin kekonvergenan dan keunikan nilai peluang jangka panjang bila berangkat dari sembarang *state* awal. *Ergodic Markov Chain* memenuhi sifat *irreducible*, *aperiodic* dan *positive recurrent*.

1. Irreducible

Suatu rantai Markov disebut *irreducible* (tak tereduksi) jika setiap keadaan saling berkomunikasi dengan setiap keadaan lain atau dengan kata lain hanya memiliki satu kelas komunikasi. Keadaan i dan j dinamakan saling berkomunikasi (dinotasikan dengan $i \leftrightarrow j$), apabila keadaan j dapat diperoleh dari keadaan i dan keadaan i dapat diperoleh dari keadaan j .

Hal ini dapat ditunjukkan melalui diagram transisi *state* berikut:



Gambar 12. Diagram Transisi *State*

Dari gambar diatas diperlihatkan bahwa setiap *state* saling berkomunikasi, tidak ada satu atau kelompok *state* yang terisolasi di mana suatu *state* tidak dapat mencapai keadaan lain. dapat dituliskan $1 \leftrightarrow 2$, $1 \leftrightarrow 3$, $2 \leftrightarrow 3$, dengan satu kelas komunikasi yakni $\{1,2,3\}$. Maka Rantai Markov memiliki sifat *irreducible*.

2. Aperiodic

State i dapat dikatakan aperiodic jika dan hanya jika $d(i) = 1$, dimana $d(i) = \gcd\{n \geq 1, P_{ii}^n > 0\}$. Dimana $d(i)$ merupakan *greatest common divisor* atau yang biasa disebut faktor persekutuan terbesar dari semua kemungkinan n yang membuat proses dari state i kembali ke state i lagi dalam langkah positif.

$$d(1) = \gcd(1,2,3, \dots) = 1$$

$$d(2) = \gcd(1,2,3, \dots) = 1$$

$$d(3) = \gcd(1,2,3, \dots) = 1$$

Berdasarkan *greatest common divisor* (FPB) diatas, dapat ditunjukkan bahwa state 1,2,3 bersifat *aperiodic* karena $d(1) = d(2) = d(3) = 1$.

3. Positive recurrent

Ruang *states* dikatakan *recurrent* jika $f_{ii} = 1$, untuk semua state i , didefinisikan

$f_{ii} = P(X_n = i, \text{ untuk beberapa } n \geq 1 | X_0 = i)$, dengan f_{ii} artinya keadaan i dicapai dari keadaan i .

$$f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)} = f_{ii}^1 + f_{ii}^2 + f_{ii}^3 + \dots$$

Maka dari gambar 12 dapat diperlihatkan jika sistem berangkat dari state i baik dari state 1, 2 ataupun 3, kemudian bertransisi ke *state* manapun akan selalu ada cara untuk kembali ke i atau dalam hal ini $f_{ii} = 1$ maka Rantai Markov memiliki keadaan *recurrent*. Suatu *state* disebut

berulang positif (*Positive Reccurent*) jika *state* tersebut *reccurent* serta berlaku: jika proses dimulai dari *state i* maka nilai harapan dari waktu sampai proses tersebut kembali ke *state i* ($\mu_i = \frac{1}{\pi_i}$) adalah bilangan hingga (*finite*). Melalui program R ditemukan:

$$\mu_i = [4,289146 \ 2,729457 \ 2,497003]$$

Maka Rantai Markov memnuhi sifat *positive reccurent*.

Sehingga berdasarkan pembuktian di atas rantai Markov dengan matriks peluang transisi tersebut merupakan rantai Markov yang ergodik (*irreducible, aperiodic dan positive recurrent*).

Steady State dari *P* dapat diperoleh dengan menyelesaikan sistem persamaan linear. Keadaan setimbang *v* memenuhi $v = vP$ dan $\sum_i \pi_i = 1$, sehingga dapat dituliskan sistem persamaan linear sebagai berikut:

$$[\pi_1 \ \pi_2 \ \pi_3] = [\pi_1 \ \pi_2 \ \pi_3] \begin{bmatrix} 0,714 & 0,143 & 0,143 \\ 0,182 & 0,545 & 0,273 \\ 0 & 0,333 & 0,667 \end{bmatrix}$$

Lebih lanjut dijabarkan sebagai berikut:

$$\begin{aligned} \pi_1 P_{11} + \pi_2 P_{21} + \pi_3 P_{31} &= \pi_1 \\ \pi_1 P_{12} + \pi_2 P_{22} + \pi_3 P_{32} &= \pi_2 \\ \pi_1 P_{13} + \pi_2 P_{23} + \pi_3 P_{33} &= \pi_3 \\ \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}$$

Substitusi nilai peluang transisi ke dalam

$$\begin{aligned} \pi_1(0,714) + \pi_2(0,182) + \pi_3(0) &= \pi_1 \\ \pi_1(0,143) + \pi_2(0,545) + \pi_3(0,333) &= \pi_2 \\ \pi_1(0,143) + \pi_2(0,273) + \pi_3(0,667) &= \pi_3 \\ \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}$$

Dengan mengeliminasi persamaan diatas diperoleh persamaan sebagai berikut:

$$\begin{aligned} 0,286 \pi_1 - 0,182 \pi_2 &= 0 \\ 0,19 \pi_1 - 0,788 \pi_2 &= 0,333 \end{aligned}$$

Dengan menyelesaikan persamaan diatas diperoleh nilai dari $\pi_1 = 0,2331$, $\pi_2 = 0,3664$, $\pi_3 = 0,4005$

Maka dapat disusun vektor *steady state* sebagai berikut:

$$v = [0,2331 \ 0,3664 \ 0,4005]$$

Hasil yang sama diperoleh juga dengan mencari peluang jangka panjang menggunakan Persamaan Chapman Kolmogorov pada Persamaan diatas, sebagai berikut:

- Matriks peluang transisi langkah ke-2

$$\begin{aligned} P^2 &= \begin{bmatrix} 0,714 & 0,143 & 0,143 \\ 0,182 & 0,545 & 0,273 \\ 0 & 0,333 & 0,667 \end{bmatrix} \begin{bmatrix} 0,714 & 0,143 & 0,143 \\ 0,182 & 0,545 & 0,273 \\ 0 & 0,333 & 0,667 \end{bmatrix} \\ &= \begin{bmatrix} 0,535 & 0,227 & 0,236 \\ 0,229 & 0,414 & 0,359 \\ 0,06 & 0,403 & 0,536 \end{bmatrix} \end{aligned}$$

- Matriks peluang transisi langkah ke-3

$$\begin{aligned} P^3 &= \begin{bmatrix} 0,535 & 0,227 & 0,236 \\ 0,229 & 0,414 & 0,359 \\ 0,06 & 0,403 & 0,536 \end{bmatrix} \begin{bmatrix} 0,714 & 0,143 & 0,143 \\ 0,182 & 0,545 & 0,273 \\ 0 & 0,333 & 0,667 \end{bmatrix} \\ &= \begin{bmatrix} 0,423 & 0,278 & 0,296 \\ 0,238 & 0,377 & 0,386 \\ 0,116 & 0,406 & 0,476 \end{bmatrix} \\ &\vdots \end{aligned}$$

- Matriks peluang transisi langkah ke-18

$$\begin{aligned} P^{18} &= \begin{bmatrix} 0,535 & 0,227 & 0,236 \\ 0,229 & 0,414 & 0,359 \\ 0,06 & 0,403 & 0,536 \end{bmatrix} \begin{bmatrix} 0,714 & 0,143 & 0,143 \\ 0,182 & 0,545 & 0,273 \\ 0 & 0,333 & 0,667 \end{bmatrix} \\ &= \begin{bmatrix} 0,2331 & 0,3664 & 0,4005 \\ 0,2331 & 0,3664 & 0,4005 \\ 0,2331 & 0,3664 & 0,4005 \end{bmatrix} \end{aligned}$$

Atau dapat dituliskan sebagai vektor *steady state* sebagai berikut

$$v = [0,2331 \ 0,3664 \ 0,4005]$$

Matriks di atas telah mengalami keadaan stabil (ekuilibrium). Peluang transisi pada keadaan seimbang menggambarkan peluang transisi yang tidak akan berubah seiring berjalannya waktu. Terlihat bahwa ketiga kolom matriks memiliki elemen-elemen yang sama, menunjukkan bahwa probabilitas transisi sudah konvergen, yang artinya seluruh *state* pada waktu *t* dapat berpindah ke satu *state* pada waktu *t + 1* dengan probabilitas sebesar π . Untuk periode selanjutnya seperti periode ke-19, ke-20, ke-21, dan seterusnya akan dihasilkan matriks yang sama karena proses telah mencapai keadaan keseimbangan.

Dari vektor *steady state* tersebut dapat dijelaskan pula bahwa probabilitas seluruh kabupaten/kota pada *cluster-cluster* kemiskinan tinggi (ekstrem), menengah, dan rendah akan bertransisi seluruhnya menjadi kabupaten/kota pada *cluster* kemiskinan rendah dalam kondisi *Steady* atau setimbang tanpa memperhitungkan kondisi awal sebesar 40,05 %. Lebih lanjut, semua kabupaten/kota di Provinsi Sumatera Utara pada semua *cluster* akan bertransisi seluruhnya menjadi kabupaten/kota pada *cluster* kemiskinan menengah

dengan probabilitas sebesar 36,64%, juga probabilitas semua kabupaten/kota di semua *cluster* akan bertransisi seluruhnya menjadi kabupaten/kota pada *cluster* kemiskinan tinggi(ekstrem) dengan probabilitas sebesar 23,5 %.

Dari kondisi *steady* tersebut dapat juga diperlihatkan peluang Provinsi Sumatera Utara akan sepenuhnya terbebas dari *cluster* kemiskinan ekstrem, yang dalam hal ini dapat dituliskan sebagai π_1^c . Dalam keadaan setimbang dipenuhi $\sum_i \pi_i = 1$, maka dapat dituliskan:

$$\begin{aligned}\pi_1 + \pi_2 + \pi_3 &= 1 \\ \pi_2 + \pi_3 - \pi_1^c &= 0 \\ \pi_1^c = \pi_2 + \pi_3 &= 0,3664 + 0,4005 \\ &= 0,7669 = 76,69\%\end{aligned}$$

Dengan demikian model ini akan memproyeksikan bahwa pada transisi ke-18 atau pada tahun 2041, Provinsi Sumatera Utara akan terbebas dari *cluster* kemiskinan ekstrem dengan peluang sebesar 76,69 %.

PENUTUP

SIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dipaparkan pada bagian sebelumnya, dapat ditarik kesimpulan sebagai berikut:

1. Hasil *K-Means Clustering* menunjukkan bahwa dengan penentuan jumlah *cluster* optimal menggunakan *Elbow Method* baik pada data 2023 dan 2024, data akan terkelompok ke dalam 3 *cluster* dengan nilai *between cluster sum of squares* 82,7% dan 81% dimana *Cluster 1* dengan *centroid* (0,02251, 0,00561) pada tahun 2023 dan (0,01885, 0,0037) pada tahun 2024, *Cluster 2* dengan *centroid* (0,01384, 0,003263) pada tahun 2023 dan (0,01225, 0,00275) pada tahun 2024, dan *Cluster 3* dengan *centroid* (0,00833, 0,001593) pada tahun 2023 dan (0,007778, 0,001328) pada tahun 2024.
2. Keadaan setimbang atau *steady state* diperoleh pada periode ke-18 dengan vektor *steady state* sebagai berikut:

$$v = [0,2331 \quad 0,3664 \quad 0,4005]$$
3. Pada Tahun 2041 atau pada periode transisi ke-18 Provinsi Sumatera Utara akan diproyeksikan sepenuhnya bertransisi dari *cluster* kemiskinan tinggi (ekstrem) dengan probabilitas sebesar 76,69%.

SARAN

Dalam penelitian ini pengelompokan hanya difokuskan berdasarkan Indeks Kedalaman dan Keparahannya Kemiskinan. Oleh karena itu, penulis memberikan saran untuk penelitian selanjutnya untuk menggunakan fitur-fitur lain misalnya tingkat pengangguran atau indikator-indikator kesehatan seperti akses ke air bersih dan sanitasi layak, angkas stunting, dan lain-lain.

DAFTAR PUSTAKA

- Adon, M. J., & Jeraman, G. T. (2023). Kontribusi Teori Kemiskinan Sebagai Deprivasi Kapabilitas Dari Amartya Sen Dalam Upaya Pemberdayaan Masyarakat Miskin. *Jurnal Masyarakat Madani*, 8(1), 1-0. <https://doi.org/10.24014/jmm.v8i1.22295>
- Agustianto, R., Ika Purnamasari, & Suyitno. (2020). Analisis Data Ketinggian Permukaan Air Sungai Mahakam Daerah Kutai Kartanegara Tahun 2010-2016 Menggunakan Model Autoregressive Integrated Moving Average (ARIMA) Dengan Efek Outlier (Studi Kasus: Data Rata-rata Ketinggian Tiap Bulan Permukaan Air Sungai Mahakam, Tenggarong, Kalimantan Timur). *Jurnal EKSPONENSIAL*, 11(1), 39-46. <https://doi.org/10.30872/eksponensial.v11i1.643>
- Amelia, N., Oktarahmadini, S., & Harahap, A. (2025). Peran Pemerintah dalam Mengatasi Kemiskinan di Kota Medan. *Jurnal Ilmu Komunikasi, Administrasi Publik Dan Kebijakan Negara*, 2(1), 1-23. <https://doi.org/10.62383/komunikasi.v2i1.76>
- Aprilia, K., & Sembiring, F. (2021). Analisis Garis Kemiskinan Makanan Menggunakan Metode Algoritma K-Means Clustering. *SISMATIK (Seminar Nasional Sistem Informasi Dan Manajemen Informatika)*, 1-10.
- Badan Pusat Statistik. (2024a). *Data dan Informasi Kemiskinan Kabupaten/Kota di Indonesia Tahun 2024* (Vol. 16).
- Badan Pusat Statistik. (2024b). Profil Kemiskinan di Indonesia September 2024. *Berita Resmi Statistik*, 1(7), 1-6.
- BPS Provinsi Sumatera Utara. (2024). BADAN PUSAT STATISTIK PROVINSI SUMATERA UTARA. *Berita Resmi Statistik*, 41, 1-12.
- Gavira-Durón, N., Gutierrez-Vargas, O., & Cruz-Aké, S. (2021). Markov chain k-means cluster models and their use for companies' credit quality and default probability estimation. *Mathematics*, 9(8), 1-14. <https://doi.org/10.3390/math9080879>

- Habib, U., & Wahyudi, H. (2022). Indeks Kedalaman Kemiskinan Sebelum dan Saat Pandemi Covid-19 di Indonesia. *Studi Ekonomi Dan Kebijakan Publik*, 1(1), 59–72. <https://doi.org/10.35912/sekp.v1i1.1424>
- Haughton, J., & Khandker, S. R. (2009). *Handbook on Poverty and Inequality*. The World Bank.
- Huang, Z., Zheng, H., Li, C., & Che, C. (2024). Application of Machine Learning-Based K-means Clustering for Financial Fraud Detection. *Academic Journal of Science and Technology*, 10(1), 2771–3032. <https://doi.org/10.54097/74414c90>
- Kaufman, L., & Rousseeuw, P. J. (1989). *Finding Groups in Data An Introduction to Cluster Analysis*. A JOHN WILEY & SONS, INC., PUBLICATION.
- Lestari, S. S. S., & Jasuni, A. Y. (2023). Analisis Rantai Markov Lima Status pada Return Harga Saham BBKA. *Jurnal Bisnisan: Riset Bisnis Dan Manajemen*, 5(1), 108–116. <https://doi.org/10.52005/bisnisan.v5i1.138>
- Li, S., Ma, H., & Li, W. (2017). Typical solar radiation year construction using k-means clustering and discrete-time Markov chain. *Applied Energy*, 205, 720–731. <https://doi.org/10.1016/j.apenergy.2017.08.067>
- Nasib, S. K., Hasan, R., Djakaria, I., Payu, M. R. F., Nuha, A. R., & Nashar, L. O. (2024). Analisis Peluang Jangka Panjang Mesin Penggilingan Padi Menggunakan Rantai Markov. *Euler: Jurnal Ilmiah Matematika, Sains Dan Teknologi*, 12(1), 63–70. <https://doi.org/10.37905/euler.v12i1.25280>
- Noor Achmad, K., Mokhammad, M. H., Hosen, M. H., & et al. (2024). Peta Kemiskinan Ekstrem Nasional : Data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE). www.baznas.go.id;
- Noviandy, T. R., Hardi, I., Zahriah, Z., & et al. (2024). Environmental and Economic Clustering of Indonesian Provinces: Insights from K-Means Analysis. *Leuser Journal of Environmental Studies*, 2(1), 41–51. <https://doi.org/10.60084/ljes.v2i1.181>
- Olukanmi, P. O., & Twala, B. (2021). K-means-sharp: Modified centroid update for outlier-robust k-means clustering. *2021 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference, PRASA-RobMech 2021, 2018-January*, 14–19. <https://doi.org/10.1109/RoboMech.2021.8261116>
- Permatasari, A. (2024). Mengkritisi Predikat Provinsi Daerah Istimewa Yogyakarta sebagai Provinsi Termiskin di Pulau Jawa. *JAP: Jurnal Administrasi Publik*, 15(2). <https://doi.org/10.31506/jap.v15i2.28779>
- Riyono, J., Pujiastuti, C. E., & Riyana Putri, A. L. (2022). Forecasting Laju Inflasi Indonesia Menggunakan Rantai Markov. *Jurnal Sains Matematika Dan Statistika*, 8(1), 1. <https://doi.org/10.24014/jsms.v8i1.14767>
- Ross, S. M. (1996). *Stochastic processes Second Edition* (2nd ed.). University of California.
- Ross, S. M. (2010). *Introduction to Probability Models, Tenth Edition* (10th ed.). University of Southern California.
- Shiffler, R. E. (1988). Maximum z scores and outliers. *American Statistician*, 42(1), 79–80. <https://doi.org/10.1080/00031305.1988.10475530>
- Stevens, A. H. (2012). Poverty Transitions. In P. N. Jefferson (Ed.), *The Oxford Handbook of the Economics of Poverty* (pp. 494–518). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195393781.013.0016>
- Tan, S., & Syafri, M. (2024). Analisis komparatif kemiskinan di Sumatera Periode 2010-2020. *Jurnal Paradigma Ekonomika*, 19(1), 2085–1960.
- United Nation. (1995). *The Copenhagen declaration and programme of action: world summit for social development 6–12 March 1995*.
- World Bank. (2005). *Introduction to poverty analysis*. The World Bank.
- Youn, H. M., Lee, H. J., Lee, D. W., & Park, E. C. (2020). The impact of poverty transitions on frailty among older adults in South Korea: Findings from the Korean longitudinal study of ageing. *BMC Geriatrics*, 20(1). <https://doi.org/10.1186/s12877-020-01522-x>