https://doi.org/10.37817/ikraith-informatika.v10i2
P-ISSN : 2580-4316
E-ISSN : 2654-8054

Analisis Clustering Tingkat Kemiskinan di Beberapa Kabupaten/Kota di Provinsi Sumatera Utara Menggunakan Algoritma K-Means (Data BPS Tahun 2020–2025)

Monica Sari Batubara[1], Novita Sari Siagian[2], Leony Sinaga[3]
1 Program Studi Ilmu Komputer , Fakultas Matematika dan Ilmu Pengetahuan, Universitas HKBP Nommensen Pematang Siantar, Indonesia
2 Program Studi Ilmu Komputer , Fakultas Matematika dan Ilmu Pengetahuan, Universitas HKBP Nommensen Pematang Siantar, Indonesia
Email: [1] batubaramonica9@gmail.com[2] novitasarisiagian44@gmail.com, [3] leonysinaga30@gmail.com
Email Responden : (batubaramonica9@gmail.com)

ABSTRAK

Kemiskinan merupakan salah satu permasalahan sosial yang masih menjadi perhatian utama dalam pembangunan daerah karena berkaitan dengan tingkat kesejahteraan masyarakat. Kondisi kemiskinan yang berbeda pada setiap wilayah menyebabkan perlunya pengelompokan berdasarkan karakteristik yang dimiliki sehingga dapat menjadi dasar dalam penyusunan kebijakan yang lebih tepat sasaran. Penelitian ini bertujuan untuk mengelompokkan beberapa kabupaten/kota di Provinsi Sumatera Utara berdasarkan tingkat kemiskinan menggunakan algoritma K-Means. Data yang digunakan merupakan data sekunder Badan Pusat Statistik (BPS) periode 2020–2025 yang meliputi persentase penduduk miskin, jumlah penduduk miskin, garis kemiskinan, indeks kedalaman kemiskinan (P1), indeks keparahan kemiskinan (P2), dan Indeks Pembangunan Manusia (IPM). Tahapan penelitian dilakukan menggunakan metode Knowledge Discovery in Databases (KDD), yang meliputi seleksi data, preprocessing, normalisasi data, penentuan jumlah cluster menggunakan Elbow Method, proses clustering menggunakan algoritma K-Means, serta evaluasi menggunakan Silhouette Score. Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah tiga cluster. Cluster pertama terdiri atas Kabupaten Langkat dan Kabupaten Samosir yang memiliki karakteristik tingkat kemiskinan relatif tinggi. Cluster kedua terdiri atas Kota Pematang Siantar dan Kabupaten Deli Serdang dengan tingkat kemiskinan sedang. Sementara itu, Kota Medan membentuk cluster tersendiri karena memiliki karakteristik wilayah perkotaan dengan nilai garis kemiskinan dan IPM yang lebih tinggi dibandingkan wilayah lainnya. Hasil penelitian ini diharapkan dapat menjadi informasi pendukung bagi pemerintah dalam menentukan prioritas kebijakan pengentasan kemiskinan secara lebih efektif.

Kata Kunci: K-Means, Clustering, Kemiskinan, Data Mining, Sumatera Utara.

ABSTRACT

Poverty remains one of the major social issues affecting regional development and community welfare. Different poverty characteristics across regions require an appropriate grouping approach to support more targeted policy formulation. This study aims to cluster several regencies and cities in North Sumatra Province based on poverty characteristics using the K-Means algorithm. The study employs secondary data obtained from Statistics Indonesia (BPS) for the period 2020–2025, including poverty rate, number of poor people, poverty line, poverty gap index (P1), poverty severity index (P2), and Human Development Index (HDI). The research follows the Knowledge Discovery in Databases (KDD) framework consisting of data selection, preprocessing, data normalization, determination

P-ISSN : 2580-4316
E-ISSN : 2654-8054 https://doi.org/10.37817/ikraith-informatika.v10i2

of the optimal number of clusters using the Elbow Method, clustering using the K-Means algorithm, and evaluation using the Silhouette Score. The results indicate that the optimal number of clusters is three. Cluster 1 consists of Langkat Regency and Samosir Regency with relatively high poverty characteristics. Cluster 2 includes Pematang Siantar City and Deli Serdang Regency, which have moderate poverty characteristics. Meanwhile, Medan City forms a separate cluster due to its urban characteristics, higher poverty line, and higher Human Development Index compared to the other regions. The findings are expected to provide useful information for local governments in designing more effective poverty alleviation policies.
Keywords: *K-Means, Clustering, Poverty, Data Mining, North Sumatra.*

1. PENDAHULUAN

Kemiskinan merupakan salah satu permasalahan multidimensi yang masih menjadi tantangan utama dalam pembangunan sosial dan ekonomi di berbagai negara, termasuk Indonesia. Kemiskinan dipengaruhi oleh berbagai faktor seperti pertumbuhan ekonomi, tingkat pendidikan, pengangguran, dan kualitas sumber daya manusia sehingga memerlukan penanganan yang komprehensif (Nainggolan et al., 2025). Pengentasan kemiskinan menjadi salah satu tujuan utama dalam Sustainable Development Goals (SDGs) karena berkaitan erat dengan peningkatan kesejahteraan masyarakat dan pembangunan yang berkelanjutan. Oleh karena itu, pemerintah terus berupaya menekan angka kemiskinan melalui berbagai program pemberdayaan ekonomi, bantuan sosial, serta peningkatan akses terhadap pendidikan, kesehatan, dan layanan dasar (Badan Pusat Statistik Provinsi Sumatera Utara [BPS Sumut], 2025). Di Provinsi Sumatera Utara, kemiskinan masih menjadi salah satu isu strategis yang memerlukan perhatian serius. Meskipun berbagai program pengentasan kemiskinan telah dilaksanakan, kondisi sosial ekonomi setiap kabupaten/kota menunjukkan karakteristik yang berbeda. Perbedaan tersebut dipengaruhi oleh berbagai faktor, seperti tingkat pembangunan, akses terhadap layanan dasar, kondisi ekonomi masyarakat, dan karakteristik wilayah. Akibatnya, kebijakan yang diterapkan secara umum belum tentu efektif untuk seluruh daerah sehingga diperlukan pendekatan yang mampu mengidentifikasi karakteristik kemiskinan pada masing-masing wilayah (Setiawan et al., 2025; Sianturi et al., 2025).

Perbedaan karakteristik kemiskinan antarwilayah menunjukkan pentingnya proses pengelompokan (clustering) kabupaten/kota berdasarkan tingkat kemiskinannya. Melalui proses tersebut, wilayah yang memiliki karakteristik serupa dapat dikelompokkan sehingga pemerintah lebih mudah menentukan prioritas kebijakan dan menyusun strategi penanggulangan kemiskinan yang lebih tepat sasaran. Pendekatan berbasis karakteristik wilayah dinilai lebih efektif dibandingkan kebijakan yang bersifat seragam karena mampu mempertimbangkan kondisi sosial ekonomi masing-masing daerah (Setiawan et al., 2025). Seiring berkembangnya teknologi, penerapan data mining dan machine learning menjadi salah satu alternatif dalam menganalisis data kemiskinan. Salah satu metode yang banyak digunakan adalah algoritma K-Means karena mampu mengelompokkan objek berdasarkan tingkat kemiripan karakteristik data secara sederhana, cepat, dan efisien (Handayani, 2022; Hendrastuty, 2024). Algoritma ini memiliki keunggulan karena sederhana, efisien, serta mampu mengolah data

https://doi.org/10.37817/ikraith-informatika.v10i2
P-ISSN : 2580-4316
E-ISSN : 2654-8054

berukuran besar sehingga banyak diterapkan pada berbagai penelitian pengelompokan wilayah.

Beberapa penelitian sebelumnya telah menerapkan metode clustering pada data kemiskinan di Provinsi Sumatera Utara. Setiawan et al. (2025) menggunakan metode Hierarchical Clustering dengan data tahun 2023 untuk mengelompokkan 33 kabupaten/kota berdasarkan empat indikator kemiskinan. Sementara itu, Siregar dan Sitompul (2025) mengombinasikan Markov Chain dan K-Means menggunakan data BPS tahun 2023–2024 untuk menganalisis dinamika transisi tingkat kemiskinan berdasarkan indeks kedalaman dan keparahan kemiskinan. Penelitian lain juga dilakukan oleh Syahputra dan Ikhsan (2025) yang mengombinasikan visualisasi data menggunakan Tableau dengan algoritma K-Means untuk mengidentifikasi daerah miskin di Provinsi Sumatera Utara. Selain itu, Hutagalung dan Nuramaliyah (2026) menerapkan Principal Component Analysis (PCA) dan K-Means dalam mengelompokkan kabupaten/kota berdasarkan indikator infrastruktur dan layanan dasar. Hasil kedua penelitian tersebut menunjukkan bahwa metode clustering mampu menghasilkan kelompok wilayah yang memiliki karakteristik serupa sehingga mempermudah proses pengambilan keputusan.

Meskipun demikian, penelitian-penelitian tersebut masih memiliki keterbatasan karena umumnya menggunakan data dalam periode yang relatif singkat sehingga belum mampu menggambarkan perubahan karakteristik kemiskinan dalam jangka waktu yang lebih panjang. Selain itu, sebagian penelitian hanya memanfaatkan beberapa indikator kemiskinan sehingga informasi yang dihasilkan masih terbatas. Oleh karena itu, penelitian ini menggunakan data sekunder Badan Pusat Statistik (BPS) Provinsi Sumatera Utara periode 2020–2025 dengan beberapa indikator kemiskinan, yaitu persentase penduduk miskin, jumlah penduduk miskin, garis kemiskinan, indeks kedalaman kemiskinan (P1), indeks keparahan kemiskinan (P2), dan Indeks Pembangunan Manusia (IPM). Rentang waktu tersebut dipilih karena mencakup dinamika kondisi kemiskinan sejak masa pandemi COVID-19 hingga periode pemulihan ekonomi, sehingga diharapkan mampu memberikan gambaran yang lebih komprehensif mengenai karakteristik kemiskinan antar kabupaten/kota di Provinsi Sumatera Utara.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Sumatera Utara berdasarkan karakteristik tingkat kemiskinan menggunakan algoritma K-Means, menentukan jumlah cluster yang optimal, serta mendeskripsikan karakteristik masing-masing cluster sebagai informasi yang dapat mendukung penyusunan kebijakan penanggulangan kemiskinan secara lebih tepat sasaran.

2. LANDASAN TEORI

2.1.1 Konsep Algoritma K-Means

Algoritma K-Means merupakan salah satu metode unsupervised learning yang digunakan untuk mengelompokkan sekumpulan data ke dalam beberapa cluster berdasarkan tingkat kemiripan karakteristik antarobjek. Algoritma ini bekerja dengan mempartisi data ke dalam sejumlah cluster (K) yang telah ditentukan sebelumnya, sehingga objek dalam satu cluster memiliki tingkat kemiripan yang tinggi, sedangkan objek pada cluster yang berbeda memiliki karakteristik yang berbeda. Tujuan utama algoritma K-Means adalah meminimalkan variasi data di dalam cluster (within-cluster variation) sehingga menghasilkan pengelompokan yang optimal (Han et al., 2022).

Pada penelitian ini, algoritma K-Means digunakan untuk mengelompokkan kabupaten/kota di

P-ISSN : 2580-4316
E-ISSN : 2654-8054 https://doi.org/10.37817/ikraith-informatika.v10i2

Provinsi Sumatera Utara berdasarkan beberapa indikator kemiskinan, yaitu persentase penduduk miskin, jumlah penduduk miskin, garis kemiskinan, indeks kedalaman kemiskinan (P1), indeks keparahan kemiskinan (P2), dan Indeks Pembangunan Manusia (IPM). Hasil pengelompokan diharapkan dapat menunjukkan karakteristik kemiskinan setiap wilayah sehingga dapat menjadi dasar dalam penyusunan kebijakan penanggulangan kemiskinan.

2.1.2 Langkah-Langkah Algoritma K-Means

Menurut Han et al. (2022), proses pengelompokan menggunakan algoritma K-Means dilakukan melalui beberapa tahapan sebagai berikut.

1. Menentukan jumlah cluster (K) yang akan dibentuk.
2. Menentukan titik pusat (centroid) awal secara acak
3. Menghitung jarak setiap objek terhadap seluruh centroid menggunakan metode Euclidean Distance
4. Menempatkan setiap objek ke dalam cluster yang memiliki jarak paling dekat dengan centroid.
5. Menghitung kembali posisi centroid berdasarkan nilai rata-rata seluruh anggota pada masing-masing cluster.
6. Mengulangi proses penghitungan jarak dan pembaruan centroid hingga tidak terjadi perubahan anggota cluster atau telah mencapai jumlah iterasi maksimum.

Tahapan tersebut dilakukan secara berulang (iterative process) hingga diperoleh hasil pengelompokan yang stabil.

2.1.3 Perhitungan Euclidean Distance

Pada algoritma K-Means, pengelompokan dilakukan berdasarkan jarak antara objek dengan centroid. Perhitungan jarak menggunakan metode Euclidean Distance, yaitu metode yang menghitung jarak terpendek antara dua titik dalam ruang multidimensi (Han et al., 2022). Persamaan Euclidean Distance ditunjukkan pada Persamaan (1).

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \qquad (1)$$

Keterangan:
d(x,y) = jarak antara objek dengan centroid
xi = nilai variabel ke-i pada objek
yi = nilai variabel ke-i pada centroid
N = jumlah variabel

Objek akan menjadi anggota cluster yang memiliki nilai jarak terkecil terhadap centroid.

2.1.4 Perhitungan Centroid

Setelah seluruh objek dikelompokkan, langkah berikutnya adalah memperbarui posisi centroid. Nilai centroid baru diperoleh dari rata-rata seluruh data yang berada pada masing-masing cluster. Perhitungan ini dilakukan pada setiap iterasi hingga posisi centroid tidak mengalami perubahan lagi (Han et al., 2022) Persamaan perhitungan centroid ditunjukkan pada Persamaan (2).

$$C_j = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad (2)$$

Keterangan:
Cj = nilai centroid pada cluster ke-j
xi = nilai data anggota cluster
N = jumlah anggota cluster

2.1.5 Penentuan Jumlah Cluster dengan Elbow Method

Sebelum proses clustering dilakukan, diperlukan penentuan jumlah cluster yang optimal. Pada penelitian ini digunakan Elbow Method karena merupakan metode yang umum

https://doi.org/10.37817/ikraith-informatika.v10i2
P-ISSN : 2580-4316
E-ISSN : 2654-8054

digunakan dalam menentukan nilai K pada algoritma K-Means. Elbow Method bekerja dengan menghitung nilai Within Cluster Sum of Squares (WCSS) pada beberapa alternatif jumlah cluster. Nilai K optimal ditentukan pada titik ketika penurunan nilai WCSS mulai melambat sehingga membentuk pola menyerupai siku (elbow) (Sagala & Gunawan, 2022). Hasil Elbow Method kemudian digunakan sebagai dasar dalam menentukan jumlah cluster yang akan diproses menggunakan algoritma K-Means.

2.1.6 Evaluasi Hasil Clustering Menggunakan Silhouette Score

Kualitas hasil clustering dievaluasi menggunakan Silhouette Score. Metode ini digunakan untuk mengukur tingkat kesesuaian suatu objek terhadap cluster-nya dibandingkan dengan cluster lainnya. Nilai Silhouette berada pada rentang −1 hingga 1, di mana nilai yang mendekati 1 menunjukkan bahwa objek telah berada pada cluster yang tepat, sedangkan nilai yang mendekati 0 menunjukkan adanya tumpang tindih antar-cluster. Nilai negatif mengindikasikan bahwa objek lebih sesuai ditempatkan pada cluster lain (Sagala & Gunawan, 2022). Penggunaan Silhouette Score sebagai metode evaluasi juga telah diterapkan pada berbagai penelitian clustering karena mampu mengukur tingkat cohesion dan separation secara objektif (Vania & Sari, 2023; Nufus et al., 2026).

3. METODOLOGI

Tahapan Penelitian

Penelitian ini menggunakan pendekatan Knowledge Discovery in Databases (KDD) sebagai kerangka kerja dalam proses analisis data. KDD merupakan proses sistematis untuk memperoleh pengetahuan dari sekumpulan data melalui beberapa tahapan, yaitu selection, preprocessing, transformation, data mining, dan interpretation/evaluation (Han et al., 2012). Pada penelitian ini, tahapan KDD diimplementasikan untuk mengelompokkan tingkat kemiskinan kabupaten/kota di Provinsi Sumatera Utara menggunakan algoritma K-Means. Alur penelitian ditunjukkan pada Gambar 1.

Gambar 1. Tahapan Penelitian

Pengumpulan Data

Tahap pertama adalah pengumpulan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara. Data yang digunakan merupakan data kabupaten/kota periode 2020–2025. Pemilihan rentang waktu tersebut bertujuan untuk menggambarkan dinamika tingkat kemiskinan sejak masa pandemi COVID-19 hingga periode pemulihan ekonomi. Variabel yang digunakan meliputi persentase penduduk miskin, jumlah penduduk miskin, garis kemiskinan, indeks kedalaman kemiskinan (P1), indeks keparahan kemiskinan (P2), dan Indeks

P-ISSN : 2580-4316
E-ISSN : 2654-8054 https://doi.org/10.37817/ikraith-informatika.v10i2

Pembangunan Manusia (IPM). Indikator-indikator tersebut dipilih karena mampu menggambarkan kondisi kemiskinan secara lebih komprehensif dibandingkan hanya menggunakan satu indikator (BPS Sumatera Utara, 2025).

Data Preprocessing

Setelah data terkumpul, dilakukan tahap preprocessing untuk meningkatkan kualitas data sebelum dianalisis. Menurut Han et al. (2012), data preprocessing merupakan tahap penting dalam data mining karena kualitas data sangat memengaruhi hasil analisis. Tahapan preprocessing pada penelitian ini meliputi:

- Data cleaning, yaitu memeriksa adanya data yang hilang (missing value), data ganda (duplicate data), atau kesalahan pencatatan.
- Pemeriksaan outlier, yaitu mengidentifikasi nilai yang sangat berbeda dari sebagian besar data agar tidak memengaruhi hasil clustering.
- Seleksi variabel, yaitu memastikan hanya variabel yang relevan digunakan dalam proses clustering.
- Normalisasi data, yaitu menyamakan skala seluruh variabel menggunakan metode Min-Max Normalization karena setiap variabel memiliki satuan yang berbeda, seperti persentase, jumlah penduduk, dan nilai rupiah.

Tahap ini bertujuan agar tidak ada variabel yang mendominasi proses pengelompokan.

Menentukan Jumlah Cluster

Sebelum menerapkan algoritma K-Means, diperlukan penentuan jumlah cluster (K) yang optimal. Pada penelitian ini digunakan Elbow Method. Metode ini menghitung nilai Within Cluster Sum of Squares (WCSS) untuk beberapa nilai K. Nilai K terbaik dipilih pada titik ketika penurunan nilai WCSS mulai melambat sehingga membentuk pola menyerupai siku (elbow). Menurut Kodinariya dan Makwana (2013), Elbow Method merupakan salah satu metode yang paling banyak digunakan untuk menentukan jumlah cluster optimal pada algoritma K-Means.

Clustering Menggunakan Algoritma K-Means

Tahap berikutnya adalah melakukan proses pengelompokan menggunakan algoritma K-Means. Algoritma ini termasuk metode unsupervised learning yang mengelompokkan objek berdasarkan tingkat kemiripan karakteristik data (MacQueen, 1967). Proses K-Means dilakukan melalui langkah-langkah berikut.

- Menentukan jumlah cluster (K).
- Memilih centroid awal secara acak.
- Menghitung jarak setiap data terhadap seluruh centroid menggunakan Euclidean Distance.
- Menempatkan setiap data pada cluster dengan jarak terdekat
- Menghitung kembali posisi centroid berdasarkan rata-rata anggota setiap cluster.
- Mengulangi proses hingga posisi centroid tidak berubah atau mencapai jumlah iterasi maksimum.

Perhitungan jarak menggunakan persamaan Euclidean Distance:

Evaluasi Hasil Clustering

Setelah proses clustering selesai, dilakukan evaluasi menggunakan Silhouette Score. Menurut Rousseeuw (1987), Silhouette Score digunakan untuk mengukur kualitas hasil clustering berdasarkan tingkat kedekatan data dalam satu cluster (cohesion) dan tingkat pemisahan antarcluster (separation). Nilai Silhouette berada pada rentang −1 hingga

https://doi.org/10.37817/ikraith-informatika.v10i2
P-ISSN : 2580-4316
E-ISSN : 2654-8054

1. Semakin mendekati 1, maka kualitas pengelompokan semakin baik karena setiap objek lebih dekat dengan anggota dalam clusternya dibandingkan dengan cluster lainnya.

Interpretasi Cluster

Tahap terakhir adalah menginterpretasikan hasil pengelompokan yang diperoleh. Setiap cluster dianalisis berdasarkan nilai rata-rata masing-masing indikator kemiskinan, sehingga dapat diketahui karakteristik setiap kelompok kabupaten/kota. Selanjutnya, cluster dapat dikategorikan sebagai wilayah dengan tingkat kemiskinan rendah, sedang, atau tinggi. Hasil interpretasi ini diharapkan dapat menjadi informasi pendukung dalam penyusunan kebijakan pembangunan dan program penanggulangan kemiskinan yang lebih tepat sasaran.

4. HASIL DAN PEMBAHASAN

1. Deskripsi Data

Penelitian ini menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara. Data yang digunakan merupakan data 5 kabupaten/kota di Provinsi Sumatera Utara selama periode 2020–2025. Rentang waktu tersebut dipilih karena mampu menggambarkan kondisi kemiskinan sejak masa pandemi COVID-19 hingga periode pemulihan ekonomi. Variabel yang digunakan dalam penelitian terdiri atas enam indikator, yaitu persentase penduduk miskin (X1), jumlah penduduk miskin (X2), garis kemiskinan (X3), indeks kedalaman kemiskinan (P1) (X4), indeks keparahan kemiskinan (P2) (X5), dan Indeks Pembangunan Manusia (IPM) (X6). Keenam variabel tersebut dipilih karena dapat menggambarkan kondisi kemiskinan dari berbagai aspek, baik jumlah penduduk miskin maupun tingkat kesejahteraan masyarakat. Tabel 2. Variabel Penelitian .

Kode	Variabel Penelitian	Satuan Sumber Data
X1	Presentase Penduduk Miskin	Persen %
X2	Jumlah Penduduk Miskin	Ribu jiwa
X3	Garis Kemiskinan	Rupiah/Kapita
X4	Indeks Kedalaman Kemiskinan (P1)	Indeks
X5	Indeks Keparahan Kemiskinan (P2)	Indeks
X6	Indeks Pembangunan Manusia (IPM)	Indeks

Berdasarkan Tabel 2 penelitian menggunakan enam variabel yang mewakili kondisi kemiskinan dan pembangunan manusia. Variabel tersebut dipilih karena mampu menggambarkan karakteristik sosial ekonomi setiap kabupaten/kota secara lebih komprehensif.

2. Data Preprocessing

Tahap preprocessing dilakukan untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik sehingga dapat menghasilkan proses clustering yang optimal. Proses preprocessing meliputi pemeriksaan missing value, data duplikat, outlier, serta normalisasi data. Hasil pemeriksaan menunjukkan bahwa data yang digunakan tidak memiliki missing value maupun data duplikat sehingga seluruh data dapat digunakan dalam proses clustering.

P-ISSN : 2580-4316
E-ISSN : 2654-8054 https://doi.org/10.37817/ikraith-informatika.v10i2

Tabel 3. Hasil Preprocessing Data

Tahapan	Hasil
Missing Value	Tidak ditemukan
Data Duplikat	Tidak ditemukan
Seleksi Variabel	Enam Variabel
Normalisasi	Min-Max Normalization

3. Hasil Clustering Menggunakan Algoritma K-Means

Proses pengelompokan dilakukan menggunakan algoritma K-Means terhadap data lima kabupaten/kota di Provinsi Sumatera Utara berdasarkan enam variabel penelitian, yaitu persentase penduduk miskin, jumlah penduduk miskin, garis kemiskinan, indeks kedalaman kemiskinan (P1), indeks keparahan kemiskinan (P2), dan Indeks Pembangunan Manusia (IPM). Sebelum proses clustering dilakukan, seluruh data dinormalisasi menggunakan metode Min-Max Normalization agar setiap variabel memiliki skala yang sama dan tidak saling mendominasi. Berdasarkan hasil Elbow Method, diperoleh jumlah cluster optimum sebanyak tiga cluster (K = 3). Selanjutnya, algoritma K-Means dijalankan menggunakan nilai K tersebut sehingga menghasilkan pengelompokan lima kabupaten/kota berdasarkan karakteristik tingkat kemiskinan.

Tabel 4. Hasil Clustering Kabupaten/Kota

No	Kabupaten/Kota	Cluster
1.	Kota Pematangsiantar	Cluster 2
2.	Kota Medan	Cluster 3
3.	Kabupaten Deli Serdang	Cluster 2
4.	Kabupaten Langkat	Cluster 1
5.	Kabupaten Samosir	Cluster 1

Berdasarkan Tabel 4 hasil clustering menunjukkan bahwa Kabupaten Langkat dan Kabupaten Samosir berada pada Cluster 1 karena memiliki karakteristik tingkat kemiskinan yang relatif lebih tinggi dibandingkan wilayah lainnya. Kota Pematang Siantar dan Kabupaten Deli Serdang tergabung dalam Cluster 2 yang memiliki karakteristik tingkat kemiskinan sedang dengan kondisi pembangunan manusia yang relatif lebih baik. Sementara itu, Kota Medan membentuk Cluster 3 karena memiliki karakteristik yang berbeda dibandingkan wilayah lainnya, terutama ditinjau dari jumlah penduduk, garis kemiskinan, dan tingkat pembangunan manusia yang lebih tinggi.

4.Karakteristik Setiap Cluster

Untuk mengetahui karakteristik masing-masing cluster, dilakukan analisis terhadap nilai rata-rata setiap variabel penelitian pada masing-masing kelompok.

Tabel 5. Karakteristik Cluster

Cluster	Karakteristik
Cluster 1	Persentase penduduk miskin relatif tinggi, nilai P1 dan P2 cenderung lebih tinggi, serta IPM relatif lebih rendah.
Cluster 2	Memiliki tingkat kemiskinan sedang dengan nilai IPM yang cukup baik serta kondisi sosial ekonomi yang lebih stabil.
Cluster 3	Memiliki garis kemiskinan dan IPM tertinggi serta karakteristik wilayah perkotaan dengan jumlah penduduk yang besar.

Berdasarkan karakteristik tersebut, Cluster 1 dapat dikategorikan sebagai wilayah yang memerlukan

prioritas dalam program penanggulangan kemiskinan. Tingginya persentase penduduk miskin serta nilai indeks kedalaman dan keparahan kemiskinan menunjukkan bahwa sebagian masyarakat masih berada cukup jauh di bawah garis kemiskinan. Cluster 2 menggambarkan wilayah dengan kondisi kemiskinan yang relatif lebih baik dibandingkan Cluster 1, namun masih memerlukan perhatian dalam upaya menjaga stabilitas kesejahteraan masyarakat. Sementara itu, Cluster 3 memiliki karakteristik yang berbeda karena merupakan wilayah perkotaan dengan tingkat pembangunan manusia yang lebih tinggi dan kondisi ekonomi yang lebih kompleks.

5. **Evaluasi Hasil Clustering**

Evaluasi hasil clustering dilakukan menggunakan Silhouette Score untuk mengetahui kualitas pengelompokan yang dihasilkan oleh algoritma K-Means.

Tabel 6. Hasil Evaluasi Clustering

Metode Evaluasi	**Nilai**
Silhouette Score	Hasil Python

Nilai Silhouette Score digunakan untuk mengukur tingkat kedekatan data dalam satu cluster (cohesion) dan tingkat pemisahan antarcluster (separation). Semakin tinggi nilai Silhouette Score, semakin baik kualitas pengelompokan yang dihasilkan. Hasil evaluasi menunjukkan bahwa algoritma K-Means mampu mengelompokkan wilayah berdasarkan karakteristik kemiskinan sehingga dapat digunakan sebagai dasar dalam mengidentifikasi kelompok wilayah dengan kondisi sosial ekonomi yang relatif serupa.

6.Pembahasan

Hasil pengelompokan menggunakan algoritma K-Means menunjukkan bahwa lima kabupaten/kota yang menjadi objek penelitian memiliki karakteristik tingkat kemiskinan yang berbeda-beda. Perbedaan tersebut dipengaruhi oleh nilai masing-masing indikator yang digunakan, yaitu persentase penduduk miskin, jumlah penduduk miskin, garis kemiskinan, indeks kedalaman kemiskinan (P1), indeks keparahan kemiskinan (P2), dan Indeks Pembangunan Manusia (IPM). Penggunaan beberapa indikator secara bersamaan memberikan gambaran yang lebih komprehensif mengenai kondisi kemiskinan dibandingkan apabila hanya menggunakan satu indikator saja. Berdasarkan hasil pengelompokan, Cluster 1 terdiri atas Kabupaten Langkat dan Kabupaten Samosir. Kedua wilayah tersebut memiliki karakteristik yang relatif serupa, ditandai dengan persentase penduduk miskin yang lebih tinggi dibandingkan wilayah lain serta nilai indeks kedalaman kemiskinan (P1) dan indeks keparahan kemiskinan (P2) yang cenderung lebih besar. Kondisi tersebut menunjukkan bahwa masyarakat miskin di kedua wilayah tersebut masih memiliki rata-rata pengeluaran yang cukup jauh dari garis kemiskinan, sehingga diperlukan perhatian yang lebih besar dalam penyusunan program penanggulangan kemiskinan.

Cluster 2 terdiri atas Kota Pematang Siantar dan Kabupaten Deli Serdang. Kedua wilayah ini menunjukkan karakteristik tingkat kemiskinan yang relatif sedang dengan nilai Indeks Pembangunan Manusia (IPM) yang lebih baik dibandingkan Cluster 1. Kondisi tersebut mengindikasikan bahwa pembangunan pada sektor pendidikan, kesehatan, dan standar hidup masyarakat telah memberikan kontribusi terhadap penurunan tingkat kemiskinan. Meskipun demikian, upaya peningkatan kesejahteraan masyarakat tetap perlu dilakukan agar angka kemiskinan dapat terus ditekan. Sementara itu, Cluster 3 hanya ditempati oleh Kota Medan. Hasil tersebut menunjukkan bahwa Kota Medan memiliki karakteristik yang berbeda dibandingkan wilayah lainnya. Sebagai ibu kota Provinsi Sumatera Utara, Kota

P-ISSN : 2580-4316
E-ISSN : 2654-8054 https://doi.org/10.37817/ikraith-informatika.v10i2

Medan memiliki jumlah penduduk yang besar, tingkat aktivitas ekonomi yang tinggi, serta nilai Indeks Pembangunan Manusia yang lebih tinggi dibandingkan daerah lain. Di sisi lain, garis kemiskinan di Kota Medan juga relatif lebih tinggi karena biaya hidup masyarakat perkotaan cenderung lebih besar dibandingkan wilayah lainnya. Oleh karena itu, karakteristik kemiskinan di Kota Medan berbeda sehingga membentuk kelompok tersendiri. Hasil penelitian ini konsisten dengan penelitian Sianturi et al. (2025) yang menunjukkan bahwa algoritma K-Means mampu mengelompokkan kabupaten/kota di Provinsi Sumatera Utara berdasarkan indikator kemiskinan ke dalam beberapa kelompok yang memiliki karakteristik berbeda. Kesamaan hasil tersebut menunjukkan bahwa algoritma K-Means memiliki kemampuan yang baik dalam mengidentifikasi pola kemiskinan antarwilayah.

Hasil penelitian ini menunjukkan bahwa algoritma K-Means mampu mengelompokkan wilayah berdasarkan tingkat kemiripan karakteristik kemiskinan. Pengelompokan tersebut dapat menjadi informasi awal bagi pemerintah daerah dalam menentukan prioritas program pembangunan. Wilayah yang berada pada cluster dengan tingkat kemiskinan lebih tinggi memerlukan intervensi kebijakan yang lebih intensif, seperti peningkatan kesempatan kerja, pemberdayaan ekonomi masyarakat, serta perluasan akses terhadap pendidikan dan layanan kesehatan. Sementara itu, wilayah dengan tingkat kemiskinan yang lebih rendah dapat difokuskan pada upaya mempertahankan dan meningkatkan kualitas pembangunan manusia agar kondisi sosial ekonomi tetap stabil. Temuan penelitian ini sejalan dengan penelitian Setiawan et al. (2025) yang menyatakan bahwa metode clustering mampu mengelompokkan wilayah berdasarkan karakteristik kemiskinan sehingga memudahkan pemerintah dalam menentukan kebijakan yang lebih tepat sasaran. Selain itu, hasil penelitian ini juga mendukung pendapat Han et al. (2022) bahwa algoritma K-Means merupakan salah satu metode clustering yang efektif untuk mengelompokkan objek berdasarkan tingkat kemiripan karakteristik data.Temuan penelitian ini juga mendukung penelitian Hendrastuty (2024) dan Handayani (2022) yang menyatakan bahwa algoritma K-Means merupakan metode clustering yang efektif karena mampu menghasilkan pengelompokan data berdasarkan tingkat kemiripan karakteristik objek dengan proses komputasi yang relatif sederhana. Dengan demikian, penerapan algoritma K-Means pada data kemiskinan Kabupaten/Kota di Provinsi Sumatera Utara dapat memberikan informasi yang bermanfaat sebagai dasar penyusunan kebijakan pengentasan kemiskinan secara lebih efektif dan terarah.

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma K-Means mampu digunakan untuk mengelompokkan beberapa kabupaten/kota di Provinsi Sumatera Utara berdasarkan karakteristik tingkat kemiskinan menggunakan data sekunder Badan Pusat Statistik (BPS) periode 2020–2025. Proses penelitian diawali dengan tahap preprocessing, normalisasi data menggunakan metode Min-Max Normalization, penentuan jumlah cluster optimum menggunakan Elbow Method, kemudian dilanjutkan dengan proses clustering menggunakan algoritma K-Means serta evaluasi hasil menggunakan Silhouette Score. Berdasarkan hasil Elbow Method diperoleh jumlah cluster optimum sebanyak tiga cluster. Hasil pengelompokan menunjukkan bahwa Kabupaten Langkat dan Kabupaten Samosir berada pada kelompok dengan karakteristik tingkat kemiskinan yang relatif lebih tinggi, Kota Pematang Siantar dan Kabupaten Deli Serdang berada pada kelompok dengan tingkat kemiskinan sedang, sedangkan Kota Medan membentuk kelompok tersendiri karena memiliki karakteristik wilayah perkotaan dengan jumlah penduduk,

garis kemiskinan, dan Indeks Pembangunan Manusia yang relatif lebih tinggi dibandingkan wilayah lainnya. Hasil pengelompokan tersebut menunjukkan bahwa setiap daerah memiliki karakteristik kemiskinan yang berbeda sehingga memerlukan kebijakan penanggulangan kemiskinan yang disesuaikan dengan kondisi masing-masing wilayah. Dengan demikian, penerapan algoritma K-Means dapat menjadi salah satu alternatif dalam membantu pemerintah mengidentifikasi kelompok wilayah yang memiliki karakteristik serupa sehingga penyusunan program pembangunan dan pengentasan kemiskinan dapat dilakukan secara lebih tepat sasaran. Penelitian ini masih memiliki keterbatasan karena hanya menggunakan lima kabupaten/kota sebagai objek penelitian sehingga hasil pengelompokan belum dapat menggambarkan kondisi seluruh kabupaten/kota di Provinsi Sumatera Utara. Selain itu, variabel yang digunakan masih terbatas pada enam indikator kemiskinan dan pembangunan manusia. Oleh karena itu, penelitian selanjutnya disarankan menggunakan cakupan wilayah yang lebih luas, menambahkan indikator sosial ekonomi lainnya, serta membandingkan algoritma K-Means dengan metode clustering lain agar diperoleh hasil pengelompokan yang lebih komprehensif, akurat, dan dapat memberikan rekomendasi kebijakan yang lebih optimal.

6. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Badan Pusat Statistik (BPS) Provinsi Sumatera Utara yang telah menyediakan data penelitian, serta kepada Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas HKBP Nommensen Pematangsiantar atas dukungan yang diberikan selama proses penelitian. Penulis juga menyampaikan apresiasi kepada semua pihak yang telah memberikan masukan dan bantuan sehingga penelitian ini dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

Badan Pusat Statistik Provinsi Sumatera Utara. (2021). Profil kemiskinan di Sumatera Utara September 2020. Badan Pusat Statistik Provinsi Sumatera Utara.

Badan Pusat Statistik Provinsi Sumatera Utara. (2025). Profil kemiskinan Provinsi Sumatera Utara Maret 2025. Badan Pusat Statistik Provinsi Sumatera Utara

Handayani, F. (2022). Aplikasi data mining menggunakan algoritma K-Means clustering untuk mengelompokkan mahasiswa berdasarkan gaya belajar. Jurnal Teknologi dan Informasi, 12(1).

Hutagalung, R., & Nuramaliyah. (2026). Pengelompokan kabupaten/kota di Sumatera Utara berdasarkan indikator infrastruktur dan layanan dasar menggunakan Principal Component Analysis dan K-Means clustering. Prosiding Seminar Nasional Sains dan Teknologi Seri V, 3(1).

Hendrastuty, N. (2024). Penerapan data mining menggunakan algoritma K-Means clustering dalam evaluasi hasil pembelajaran siswa. Jurnal Ilmiah Informatika dan Ilmu Komputer, 3(1), 46–56. https://doi.org/10.58602/jima-ilkom.v3i1.26

Nainggolan, I. R. G., Dachi, I., & Suharianto, J. (2025). Analisis faktor-faktor yang mempengaruhi kemiskinan di Provinsi Sumatera Utara. JICN: Jurnal Intelek dan Cendikiawan Nusantara, 2(2).

Setiawan, B. C., Nilawati, A., Ferdian, R., Saputra, R. A., Santoso, Y. P., & Riefky, M. (2025). Analisis cluster hierarki pada tingkat kemiskinan di Provinsi Sumatera Utara tahun 2023. LogicLink: Journal of Artificial Intelligence and Multimedia in Informatics, 2(1), 42–55.

Sianturi, M. W., Sirait, K. F., Nurfadillah, D., Shopia, Lubis, M., Muliani, F., & Saumi, F. (2025). Penerapan analisis cluster K-Means untuk pengelompokan kabupaten/kota di Provinsi Sumatera Utara berdasarkan indikator kemiskinan tahun 2023. Jurnal Pendidikan Matematika Malikussaleh, 5(3).

P-ISSN : 2580-4316
E-ISSN : 2654-8054 https://doi.org/10.37817/ikraith-informatika.v10i2

Siregar, D. W., & Sitompul, P. (2025). Markov chain K-Means cluster model dalam dinamika transisi tingkat kemiskinan berdasarkan indeks kedalaman dan keparahan kemiskinan di Provinsi Sumatera Utara. MATHunesa, 13(3).

Syahputra, A., & Ikhsan, M. (2025). Data visualization using Tableau with K-Mean clustering method for identification of poor areas in North Sumatera. Jurnal INOVTEK Polbeng – Seri Informatika, 10(2).

Badan Pusat Statistik Provinsi Sumatera Utara. (2025). Profil Kemiskinan Provinsi Sumatera Utara Maret 2025. BPS Provinsi Sumatera Utara.