

PENERAPAN DATA MINING UNTUK CLUSTERING WILAYAH BERDASARKAN TINGKAT KEMISKINAN DI ACEH MENGGUNAKAN METODE K-MEANS CLUSTERING DAN DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE (DBSCAN)

Puput Ahyawan Pasaribu, Veri Ilhadi, Rahma Fitria

Sistem Informasi, Universitas Malikussaleh

Muara Satu, Lhoksemawe, Aceh, Indonesia

veri@unimal.ac.id

ABSTRAK

Kemiskinan masih menjadi permasalahan kompleks di Indonesia, yang tidak hanya berdampak pada aspek ekonomi, tetapi juga sosial dan pembangunan manusia. Salah satu provinsi yang mengalami tantangan signifikan dalam hal ini adalah Aceh, yang dalam beberapa tahun terakhir menunjukkan angka kemiskinan yang relatif tinggi dibandingkan provinsi di Sumatera. Penelitian ini bertujuan untuk mengidentifikasi dan mengelompokkan wilayah di Aceh berdasarkan tingkat kemiskinan menggunakan teknik *data mining*. Metode yang digunakan adalah *K-Means Clustering* dan *Density-Based Spatial Clustering Of Applications With Noise* (DBSCAN). Data yang dianalisis mencakup periode tahun 2020 hingga 2024, dengan variabel persentase penduduk miskin, tingkat pengangguran terbuka, dan PDRB per kapita. Hasil pengelompokan menggunakan metode *K-Means* menghasilkan 3 cluster utama, yaitu wilayah dengan kategori kemiskinan tinggi, sedang, dan rendah. Sementara itu, metode DBSCAN mampu mengidentifikasi pola distribusi data yang lebih kompleks serta mendeteksi *outlier* atau wilayah-wilayah dengan karakteristik ekstrem, yang ditandai dengan *cluster -1* (*noise*). Hasil evaluasi menggunakan *Silhouette Score*, menunjukkan bahwa DBSCAN secara umum memiliki kualitas *clustering* yang lebih baik dibandingkan *K-Means*. Meskipun pada tahun 2020 *K-Means* menunjukkan performa sedikit lebih unggul dengan nilai 0.2838, DBSCAN secara konsisten menunjukkan performa yang lebih baik dari tahun 2021 hingga 2024, dengan nilai tertinggi sebesar 0.4711 pada tahun 2022.

Kata kunci : Kemiskinan, Aceh, Data Mining, K-Means, DBSCAN, Silhouette Score

1. PENDAHULUAN

Kemiskinan merupakan salah satu permasalahan sosial dan ekonomi yang kompleks serta menjadi perhatian utama dalam pembangunan di Indonesia. Berdasarkan data Badan Pusat Statistik (BPS), tingkat kemiskinan di Indonesia tahun 2024 mencapai 9,03% dari total penduduk atau sekitar 25,9 juta jiwa. Terdapat banyak faktor yang mempengaruhi tingkat kemiskinan di Indonesia secara langsung maupun tidak langsung meliputi beberapa aspek utama berdasarkan hasil penelitian dengan analisis jalur. Faktor langsung yang signifikan mempengaruhi kemiskinan adalah tingkat pengangguran. Pengangguran berpengaruh positif signifikan terhadap kemiskinan karena semakin banyaknya pengangguran menyebabkan berkurangnya pendapatan masyarakat sehingga meningkatkan jumlah penduduk miskin [1]. Selain itu, pendapatan asli daerah (PAD) memiliki pengaruh negatif signifikan terhadap kemiskinan, artinya peningkatan PAD dapat menurunkan angka kemiskinan secara langsung. Faktor-faktor tidak langsung yang turut mempengaruhi kemiskinan antara lain pertumbuhan ekonomi dan pengeluaran pemerintah di sektor pendidikan dan kesehatan, yang berperan sebagai variabel mediasi dalam mengurangi kemiskinan [2]. Selain itu, distribusi kemiskinan tidak merata di seluruh wilayah, dengan beberapa daerah memiliki tingkat kemiskinan yang jauh lebih tinggi

dibandingkan rata-rata nasional, seperti di wilayah timur Indonesia dan beberapa provinsi lainnya.

Provinsi Aceh masih menghadapi tingkat kemiskinan yang relatif tinggi dibandingkan dengan provinsi lain di Indonesia. Berdasarkan data dari Badan Pusat Statistika tahun 2024, terdapat 804,53 ribu jiwa jumlah penduduk miskin di Provinsi Aceh atau jika dipresentasikan tercatat sebesar 14,23%. Angka ini menunjukkan penurunan sebesar 0,22 poin dibandingkan tahun 2023 yang tercatat sebesar 806,75 ribu jiwa atau 14,45%. Meskipun mengalami penurunan, kondisi ini tetap menjadi perhatian karena Provinsi Aceh memiliki tingkat kemiskinan yang lebih tinggi dibandingkan rata-rata nasional [3].

Masalah lain yang turut memperparah kondisi kemiskinan di Aceh adalah adanya ketimpangan kemiskinan yang signifikan antara wilayah perkotaan dan perdesaan di Aceh, di mana kemiskinan di perdesaan mencapai 16,75%, jauh lebih tinggi dibandingkan 9,60% di perkotaan [3]. Ketimpangan tingkat kemiskinan juga terjadi antar kabupaten/kota di Provinsi Aceh, dengan beberapa wilayah seperti Kabupaten Gayo Lues dan Aceh Tenggara memiliki tingkat kemiskinan yang jauh di atas rata-rata provinsi, sementara kota Banda Aceh menunjukkan tingkat kemiskinan yang relatif rendah [4]. Perbedaan ini disebabkan oleh berbagai faktor, termasuk rendahnya akses pendidikan di daerah-daerah dengan kemiskinan tinggi, keterbatasan lapangan pekerjaan,

serta infrastruktur yang belum memadai untuk mendukung aktivitas ekonomi masyarakat [5]. Selain itu, Aceh merupakan wilayah yang rawan bencana alam seperti gempa bumi dan banjir, yang secara signifikan memperburuk kondisi sosial ekonomi masyarakat, khususnya di daerah pedesaan yang rentan terhadap kerusakan fasilitas dan gangguan mata pencaharian. Kondisi tersebut menyebabkan penyaluran bantuan sosial dan program pemberdayaan masyarakat seringkali kurang tepat sasaran karena minimnya pemetaan wilayah secara detail berdasarkan tingkat kemiskinan dan karakteristik lokal, sehingga sulit untuk mengidentifikasi kelompok sasaran yang paling membutuhkan intervensi [4].

Berdasarkan penelitian tahun 2023 dengan judul "*Clustering* Pemetaan Tingkat Kemiskinan di Provinsi Jawa Barat Menggunakan Algoritma K-Means" menunjukkan bahwa algoritma K-Means dapat mengelompokkan wilayah berdasarkan tingkat kemiskinan secara efektif. Hasil penelitian tersebut menunjukkan bahwa dari 27 kabupaten/kota di Jawa Barat, 12 di antaranya tergolong tidak rentan, 14 rentan, dan 1 sangat rentan terhadap kemiskinan. Evaluasi clustering menggunakan koefisien siluet menghasilkan nilai 0,55, yang mengindikasikan struktur *cluster* yang sedang dengan interpretasi yang wajar terhadap penempatan *cluster* [6]. Penelitian ini hanya menggunakan satu metode *clustering* dan belum melakukan perbandingan dengan metode lain yang lebih adaptif terhadap distribusi data yang tidak merata. Berbeda dengan penelitian ini, penelitian yang dilakukan akan membandingkan metode K-Means dengan DBSCAN untuk mengidentifikasi pola distribusi kemiskinan di Aceh, sehingga dapat menentukan metode yang lebih sesuai berdasarkan karakteristik data yang digunakan. Dengan penerapan kedua metode ini, diharapkan dapat diperoleh hasil *clustering* yang optimal dalam mengidentifikasi pola kemiskinan di Provinsi Aceh.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian yang dilakukan oleh Jilang Ramadhani, dkk. dengan judul "Perbandingan Algoritma K-Means *Clustering* dan K-Medoids Dalam Mengelompokkan Tingkat Kemiskinan Di Provinsi Riau". Penelitian ini membahas tentang pengelompokan data serta perbandingan kinerja metode K-Means *Clustering* dan K-Medoids dalam menganalisis tingkat kemiskinan di Provinsi Riau. Hasil penelitian menunjukkan bahwa K-Means menghasilkan nilai *Silhouette Coefficient* sebesar 0,387 lebih tinggi dibandingkan K-Medoids sebesar 0,289. Hal ini menunjukkan *Cluster* yang dihasilkan K-Means lebih baik dalam mengelompokkan wilayah berdasarkan tingkat kemiskinan [7]. Selanjutnya penelitian yang dilakukan oleh I Wayan Setya Adi Nugraha dengan judul "*Clustering* Pemetaan Tingkat Kemiskinan di Provinsi Jawa Barat Menggunakan

Algoritma K-Means" Penelitian ini menerapkan metode *clustering* menggunakan algoritma k-means untuk data kemiskinan di provinsi Jawa Barat pada tahun 2015 sampai dengan 2020. Hasil penelitian menunjukkan bahwa terdapat jumlah anggota *cluster* 1 yaitu 12 kabupaten/kota yang dikategorikan daerah tidak rawan kemiskinan, *cluster* 2 yaitu 14 kabupaten/kota yang dikategorikan daerah sangat rawan kemiskinan [6].

Penelitian yang dilakukan oleh Talitha Naifa Audrey, dkk. dengan judul "Penerapan Algoritma DBSCAN Untuk *Clustering* Penjualan Supermarket" penelitian ini menggunakan dataset yang berisi 100 data penjualan supermarket, dengan mengambil 2 atribut. Hasil penelitian menunjukkan dengan menggunakan nilai parameter $Eps = 6$ dan nilai $MinPts = 9$ data produk barang terbagi menjadi 3 *cluster*, Dimana *cluster* 1 beranggotakan produk yang tidak diminati atau jarang dibeli *customer*, *Cluster* 2 beranggotakan data barang yang diminati *customer*, dan *cluster* 3 beranggotakan produk yang sangat diminati [8].

2.2. Data Mining

Data mining adalah proses menemukan pola-pola baru dari kumpulan data yang sangat besar. Ini mencakup metode yang berasal dari *artificial intelligence*, *machine learning*, statistika dan sistem *database*. Dengan kata lain, data mining adalah proses pencarian data otomatis untuk mendapatkan model dari database berskala besar. Data mining berfungsi untuk menemukan informasi tersembunyi dalam database dan merupakan bagian dari proses *Knowledge Discovery in Databases* (KDD), dengan tujuan menemukan informasi atau pola yang berguna dari data [9].

2.3. Clustering

Clustering adalah salah satu teknik dalam data mining yang bertujuan untuk mengelompokkan data ke dalam sejumlah grup (*cluster* atau segmen), di mana setiap *cluster* dapat berisi beberapa data yang memiliki kemiripan. Secara umum, *clustering* digunakan untuk menemukan dan mengelompokkan data yang memiliki karakteristik yang serupa antara satu dengan lainnya. Metode ini termasuk dalam kategori *unsupervised learning*, yaitu metode yang diterapkan tanpa proses pelatihan, tanpa bantuan instruktur, dan tidak membutuhkan nilai target sebagai referensi. Dalam data mining, terdapat dua pendekatan utama dalam *clustering*, yaitu *hierarchical clustering* dan *non-hierarchical clustering* [10].

Meskipun sama dengan klasifikasi dalam hal pengorganisasian data, perbedaan utama pada clustering terletak pada kesamaan karakteristik yang menjadi dasar dalam pengelompokan [11]. *Clustering* digunakan untuk membagi data ke dalam sejumlah kelompok, di mana tiap kelompok terdiri dari objek-objek yang memiliki kemiripan berdasarkan metrik

atau ukuran tertentu. Tujuan dari analisis *cluster* adalah untuk menyusun objek atau individu ke dalam kelompok-kelompok yang masing-masing memiliki karakteristik yang berbeda antar kelompok, namun memiliki keseragaman dalam kelompok yang sama [12].

2.4. K-Means

K-Means merupakan salah satu algoritma *clustering* yang paling populer dan sering digunakan dalam proses pengolahan data. Tujuan utama dari algoritma ini adalah memisahkan data ke dalam k kelompok atau *cluster*, di mana setiap data akan dimasukkan ke dalam *cluster* yang memiliki jarak terdekat secara geometris. Proses K-Means dilakukan dengan cara berikut [13]:

- Menentukan jumlah k sebagai banyaknya *cluster* yang akan dibentuk
- Menentukan k *centroid* (pusat *cluster*) awal secara acak
- Pemilihan awal pusat *cluster* (*centroid*) dilakukan secara acak dari sejumlah objek dalam *dataset* sebanyak k . Selanjutnya, untuk menghitung *centroid* dari *cluster* ke- i pada proses iterasi berikutnya, digunakan rumus berikut ini:

$$C_{kj} = \frac{x_{1j} + x_{2j} + \dots + x_{nj}}{n} \quad (1)$$

Keterangan:

C_{kj} merupakan koordinat pusat cluster ke- k pada variabel ke- j

n menyatakan jumlah data yang tergabung dalam cluster ke- k

- Menghitung Jarak Setiap Objek ke *Centroid* Masing-Masing *Cluster*. Untuk mengetahui seberapa dekat suatu data terhadap pusat *cluster*, digunakan metode pengukuran jarak dengan *Euclidean* antara data dan setiap pusat *cluster* menggunakan rumus berikut:

$$d_{(x,y)} = \sqrt{(x_1 - y_1)^2 + \dots + (x_p - y_p)^2} \quad (2)$$

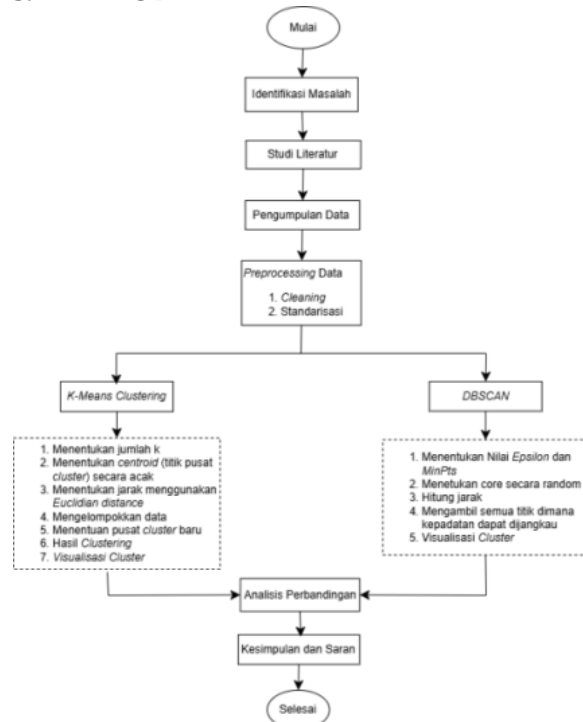
- Membandingkan hasil perhitungan jarak untuk menentukan jarak terdekat antara setiap data dan pusat *cluster*. Jarak terdekat tersebut menunjukkan bahwa data termasuk ke dalam kelompok yang sama dengan pusat *cluster* tersebut.
- Penentuan pusat *cluster* baru, dihitung berdasarkan rata-rata nilai dari seluruh data anggota *cluster*. Nilai pusat *cluster* baru ini digunakan untuk proses iterasi selanjutnya apabila hasil sebelumnya belum mencapai kondisi konvergen. Iterasi akan berhenti ketika jumlah iterasi maksimum telah tercapai atau ketika pusat *cluster* tidak lagi berubah.
- Ulangi langkah ke-3 sampai dengan langkah ke-6 hingga tidak ada lagi objek yang berpindah atau tidak terjadi perubahan pada fungsi objeknya

2.5. Density-Based Spatial Clustering Of Applications With Noise (DBSCAN)

DBSCAN adalah proses pembentukan *cluster* diidentifikasi berdasar atas kerapatan ataupun kedekatan jarak antara satu objek dengan objek lain dalam *dataset*. DBSCAN merupakan algoritma *clustering* berbasis densitas yang tidak memerlukan penentuan jumlah *cluster* di awal dan mampu mengidentifikasi *cluster* dengan berbagai bentuk dan ukuran serta mengabaikan *noise* atau data *outlier* yang tidak relevan [14]. Metode ini secara umum dilakukan dengan proses berikut [15]:

- Menentukan nilai *epsilon* (ϵ) dan *MinPts* (Jumlah minimal data dalam radius)
 - Menentukan titik awal (*core*) secara acak.
 - Melakukan pengukuran jarak titik awal dengan setiap titik dalam *dataset*, menggunakan persamaan jarak menggunakan persamaan berikut
- $$\text{Jarak} = \sqrt{(x - xp)^2 + (y - yp)^2} \quad (3)$$
- Keterangan:
- x : Nilai atribut x titik tujuan
 - y : Nilai atribut y titik tujuan
 - xp : Nilai pusat atribut x
 - yp : Nilai pusat atribut y
- Mengambil seluruh titik yang dapat dicapai berdasarkan kepadatan (*density reachable*) dari titik awal, asalkan memenuhi parameter nilai Eps dan $MinPts$. Jika titik awal tersebut adalah *core point*, maka proses pembentukan *cluster* dapat dilakukan.
 - Mengulangi proses pada langkah ketiga dan keempat hingga semua data telah dievaluasi.

3. METODE PENELITIAN



Gambar 1. Tahapan penelitian

Penelitian ini memfokuskan pada pengukuran dan analisis data berupa angka dan statistik untuk mendapatkan pemahaman mendalam terkait fenomena yang menjadi objek penelitian. Analisis data penelitian ini dilakukan menggunakan perangkat lunak *Python*. Pendekatan yang digunakan dalam penelitian ini adalah *clustering*, dengan membandingkan performa algoritma K-Means Clustering dan *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN). K-Means dipilih karena efisiensinya dalam membagi data ke dalam cluster berdasarkan *centroid*, sedangkan DBSCAN dipilih karena kemampuannya dalam mengidentifikasi *noise* serta mengelompokkan data dengan distribusi yang tidak beraturan.

3.1. Identifikasi Masalah

Pada tahap ini, dilakukan identifikasi masalah penelitian yang akan dikaji. Masalah tersebut akan dianalisis untuk memperoleh informasi yang dapat digunakan dalam penyelesaian penelitian.

3.2. Studi Literatur

Pada tahap ini, dilakukan studi literatur untuk memahami konsep, metode, dan penelitian terdahulu yang relevan dengan objek penelitian. Studi ini mencakup penelusuran referensi terkait teori, studi kasus, dan metode yang digunakan. Referensi yang digunakan bersumber dari buku, internet, website dan media lainnya.

3.3. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data yang berkaitan dengan pokok pembahasan penelitian. Data yang dikumpulkan meliputi informasi dari instansi terkait dan berbagai sumber referensi seperti buku, jurnal dan media lainnya.

3.4. Preprocessing Data

Preprocessing data dilakukan untuk membersihkan, mengubah, dan normalisasi data untuk memastikan hasil yang akurat dan relevan. Tahapan pemrosesan data meliputi:

- Data Cleaning*: Menghilangkan data yang tidak lengkap, duplikat, atau mengandung anomali.
- Standarisasi Data: Mengubah nilai-nilai dalam suatu variabel atau fitur sehingga mereka memiliki skala yang seragam atau dapat dibandingkan.

3.5. K-Means Clustering

Pada tahap ini dilakukan pengolahan data menggunakan metode K-Means Clustering. Algoritma ini bekerja dengan menentukan jumlah *cluster* awal (*k*) dan mengelompokkan data berdasarkan jarak *Euclidean* terhadap *centroid* dari setiap *cluster*. Algoritma ini digunakan untuk mengidentifikasi pola distribusi kemiskinan di wilayah yang memiliki karakteristik yang sama.

3.6. Density-Based Spatial Clustering Of Applications With Noise (DBSCAN)

Pada tahap ini dilakukan pengolahan data menggunakan metode *Density-Based Spatial Clustering Of Applications With Noise* (DBSCAN) untuk memperoleh hasil penelitian yang berupa pengelompokan wilayah di Provinsi Aceh berdasarkan tingkat kemiskinan.

3.7. Analisis Perbandingan

Hasil *Clustering* wilayah berdasarkan tingkat kemiskinan menggunakan metode K-Means *Clustering* dan *Density-Based Spatial Clustering Of Applications With Noise* (DBSCAN) akan dianalisis dan dibandingkan. Analisis ini bertujuan untuk melihat sejauh mana akurasi kedua metode yang digunakan dan bagaimana tingkat performa dari metode K-Means *Clustering* dan *Density-Based Spatial Clustering Of Applications With Noise* (DBSCAN).

4. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini mencakup seluruh wilayah di Provinsi Aceh yang terdiri dari 23 kabupaten/kota tahun 2020-2024, dengan variabel persentase penduduk miskin, tingkat pengangguran terbuka, dan PDRB per kapita.

4.1. Preprocessing Data

Preprocessing merupakan proses penting dalam data mining yang bertujuan untuk menyiapkan data mentah agar dapat diolah secara optimal oleh algoritma. Seluruh proses *preprocessing* dilakukan menggunakan bahasa pemrograman *Python*, yang menggunakan berbagai pustaka (*library*) pengolahan data seperti *pandas*, *numpy*, dan *scikit-learn*.

a. Data Cleaning

Data diperiksa untuk memastikan tidak ada nilai yang kosong (*missing values*), *outlier*, maupun data duplikat. Dalam penelitian ini, seluruh data terkonfirmasi valid dan tidak memerlukan perlakuan khusus terhadap *missing value*.

b. Standarisasi Data

Proses standarisasi diterapkan menggunakan metode *Z-Score Standardization*, yaitu dengan mengubah nilai tiap fitur menjadi distribusi yang memiliki nilai rata-rata 0 dan standar deviasi 1. Proses ini dilakukan dengan memanfaatkan *library Scikit-learn* di *Python*, menggunakan fungsi *StandardScaler* yang tersedia dalam modul *sklearn.preprocessing*. Fungsi ini akan secara otomatis menghitung rata-rata dan standar deviasi setiap fitur, kemudian mengubah data sesuai dengan rumus *Z-Score*.

Berikut data kemiskinan tahun 2020-2024 yang telah melalui tahap *preprocessing* data menggunakan *Python*.

Tabel 1. *Preprocessing* Data Penduduk Miskin

Kabupaten/ Kota	Penduduk Miskin (%)				
	2020	2021	2022	2023	2024
Simeulue	0.930401	0.985704	1.060.113	0.967434	0.96392
Aceh Singkil	1.437.295	1.399.653	1.313.382	1.349.945	1.392.616
Aceh Selatan	-0.735531	-0.754081	-0.797192	-0.842499	-0.810319
Aceh Tenggara	-0.634745	-0.685089	-0.672121	-0.733654	-0.819706
Aceh Timur	-0.376852	-0.373128	-0.334429	-0.441328	-0.422302
Aceh Tengah	-0.080423	-0.130158	-0.149949	-0.133453	-0.106256
Aceh Barat	0.885937	0.934711	0.922535	0.948775	0.935757
Aceh Besar	-0.447995	-0.493113	-0.500148	-0.444438	-0.437948
Pidie	1.149.759	1.168.682	1.191.438	1.234.881	1.245.545
Bireuen	-0.67921	-0.733083	-0.772178	-0.836279	-0.785285
Aceh Utara	0.49465	0.520762	0.58797	0.569373	0.469511
Aceh Barat Daya	0.171542	0.193802	0.143968	0.193081	0.222307
Gayo Lues	1.176.437	1.183.680	1.216.452	1.247.320	1.154.799
Aceh Tamiang	-0.673281	-0.706087	-0.74091	-0.714995	-0.763381
Nagan Raya	0.696222	0.760732	0.750562	0.759074	0.729232
Aceh Jaya	-0.735531	-0.739083	-0.772178	-0.742984	-0.738348
Bener Meriah	1.048.973	1.039.698	1.066.367	1.088.718	1.117.249
Pidie Jaya	1.137.902	1.156.683	1.085.127	1.116.706	1.148.541
Banda Aceh	-2.505.214	-2.424.874	-2.454.383	-2.416.083	-2.396.807
Kota Sabang	-0.121923	-0.11216	-0.099921	-0.068146	-0.009251
Langsa	-1.455.854	-1.419.998	-1.363.138	-1.330.745	-1.339.148
Lhokseumawe	-1.349.140	-1.360.006	-1.294.349	-1.268.548	-1.304.727
Subulussalam	0.666579	0.586754	0.612984	0.497846	0.553999

Tabel 2. *Preprocessing* Data Tingkat Pengangguran

Kabupaten/ Kota	Pengangguran (%)				
	2020	2021	2022	2023	2024
Simeulue	-0.322595	-0.040331	0.12888	0.114212	0.108836
Aceh Singkil	0.784165	1.066.952	0.598041	0.654881	0.62148
Aceh Selatan	0.104926	0.273051	-0.500222	-0.497453	-0.403808
Aceh Tenggara	-0.222707	0.260516	-0.356275	-0.349998	-0.278374
Aceh Timur	0.392604	0.553006	1.232.475	1.304.776	1.335.909
Aceh Tengah	-1.289.512	-1.335.643	-0.702815	-0.666754	-0.594686
Aceh Barat	0.408586	0.536292	0.176863	0.234361	0.152465
Aceh Besar	0.536443	0.791176	1.344.434	1.381.235	1.434.075
Pidie	0.068966	0.615682	0.096892	0.152442	0.239724
Bireuen	-0.861991	-0.621132	-0.830768	-0.81967	-0.747388
Aceh Utara	0.912022	1.046.060	0.704669	0.780491	0.861441
Aceh Barat Daya	-0.937906	-0.738128	-0.873419	-0.852438	-0.801925
Gayo Lues	-1.705.046	-1.657.382	-1.630.474	-1.633.404	-1.581.798
Aceh Tamiang	0.676286	0.026524	0.86461	0.85695	0.899617
Nagan Raya	-0.466434	-0.341178	0.043578	0.054138	-0.043866
Aceh Jaya	-0.877973	-0.976299	-1.406.556	-1.507.794	-1.390.920
Bener Meriah	-1.968.751	-1.908.088	-1.678.457	-1.742.630	-1.690.871
Pidie Jaya	0.120908	-0.934514	-0.72414	-0.710444	-0.807379
Banda Aceh	1.303.583	1.309.301	1.525.701	1.304.776	1.226.836
Kota Sabang	-0.5863	-0.938693	-0.932064	-0.939819	-1.281.847
Kota Langsa	1.387.489	0.586433	1.115.185	1.140.937	1.112.309
Lhokseumawe	2.282.487	2.236.912	1.808.264	1.714.374	1.728.573
Subulussalam	0.260751	0.189483	-0.004404	0.026832	-0.098403

Tabel 3. *Preprocessing* Data PDRB Per Kapita

Kabupaten/ Kota	PDRB Per Kapita (Juta Rupiah)				
	2020	2021	2022	2023	2024
Simeulue	-0.588252	-0.619145	-0.64058	-0.613379	-0.613718
Aceh Singkil	-1.048.444	-1.012.945	-0.99532	-1.002.797	-1.006.247
Aceh Selatan	-0.648804	-0.672488	-0.70314	-0.666854	-0.678742
Aceh Tenggara	-0.722331	-0.768977	-0.81107	-0.809661	-0.804613
Aceh Timur	-0.533756	-0.521087	-0.4914	-0.532854	-0.511112
Aceh Tengah	0.264659	0.221014	0.187831	0.251643	0.301386

Kabupaten/ Kota	PDRB Per Kapita (Juta Rupiah)				
	2020	2021	2022	2023	2024
Aceh Barat	0.826058	1.236.107	1.744.967	1.655.181	1.533.854
Aceh Besar	0.138366	0.015485	-0.02185	0.014469	0.045467
Pidie	-0.568357	-0.643463	-0.6502	-0.6203	-0.601191
Bireuen	-0.11422	-0.167295	-0.19647	-0.150986	-0.131111
Aceh Utara	0.12193	0.316718	0.576943	0.351671	0.320476
Aceh Barat Daya	-0.456769	-0.542268	-0.55396	-0.524675	-0.507533
Gayo Lues	-0.360752	-0.448917	-0.51958	-0.513351	-0.517674
Aceh Tamiang	-0.49829	-0.330463	-0.19372	-0.30323	-0.291582
Nagan Raya	1.235.213	1.520.867	1.671.407	1.797.359	1.684.185
Aceh Jaya	-0.264734	-0.295162	-0.33946	-0.340976	-0.350641
Bener Meriah	-0.269924	-0.268491	-0.31059	-0.278065	-0.258176
Pidie Jaya	-0.794127	-0.867035	-0.8867	-0.849924	-0.826089
Kota Banda Aceh	3.629.593	3.452.996	3.153.608	3.195.235	3.339.008
Kota Sabang	0.524165	0.34339	0.097084	0.047812	0.034133
Kota Langsa	-0.247434	-0.269275	-0.32572	-0.300713	-0.311865
Lhokseumawe	1.358.046	1.200.022	1.063.677	1.046.205	1.007.699
Subulussalam	-0.981837	-0.879586	-0.85576	-0.851811	-0.855917

4.2. K-Means Clustering

Hasil pengelompokan setiap wilayah diklasifikasikan ke dalam salah satu dari tiga cluster yang terbentuk, yaitu *cluster 1* merupakan Wilayah dengan tingkat kemiskinan tinggi, *cluster 2* kategori wilayah dengan tingkat kemiskinan rendah dan *cluster 3* merupakan wilayah dengan tingkat kemiskinan menengah. Berikut adalah hasil *clustering* wilayah berdasarkan tingkat kemiskinan di Aceh menggunakan metode K-Means tahun 2020-2024.

Tabel 4. Hasil K-Means Clustering

Kabupaten/Kota	2020	2021	2022	2023	2024
Simeulue	1	1	1	1	1
Aceh Singkil	1	1	1	1	1
Aceh Selatan	3	1	3	3	3
Aceh Tenggara	3	1	3	3	3
Aceh Timur	3	1	3	2	3
Aceh Tengah	3	3	3	3	3
Aceh Barat	1	1	1	1	1
Aceh Besar	3	1	2	2	2
Pidie	1	1	1	1	1
Bireuen	3	3	3	3	3
Aceh Utara	1	1	1	1	1
Aceh Barat Daya	3	3	3	3	3
Gayo Lues	1	3	1	1	1
Aceh Tamiang	3	1	3	2	3
Nagan Raya	1	3	1	1	1
Aceh Jaya	3	3	3	3	3
Bener Meriah	1	3	1	1	1
Pidie Jaya	1	3	1	1	1
Banda Aceh	2	2	2	2	2
Sabang	3	3	3	3	3
Langsa	3	1	2	2	2
Lhokseumawe	2	2	2	2	2
Subulussalam	1	1	1	1	1

Hasil *clustering* wilayah menggunakan metode K-Means pada tahun 2020 menunjukkan bahwa terdapat tiga kelompok atau *cluster*. *Cluster 1* terdiri atas 10 wilayah, *cluster 2* meliputi 4 wilayah, dan *cluster 3* terdiri dari 11 wilayah. Pada tahun 2021, terjadi pergeseran distribusi *cluster*. *Cluster 1* kini

mencakup 9 wilayah. *Cluster 2* hanya mencakup 2 wilayah, yaitu Kota Banda Aceh dan Kota Lhokseumawe. Sedangkan *Cluster 3* mencakup 12 wilayah. Pada tahun 2022, *Cluster 1* terdiri dari 10 wilayah, *Cluster 2* terdiri atas 4 wilayah *Cluster 3* terdiri dari 9. Pada tahun 2023, *Cluster 1* meliputi 10 wilayah, *Cluster 2* kembali terdiri dari 4 wilayah dan *cluster 3* mencakup 9 wilayah lainnya. Pada tahun 2024, *Cluster 1* masih konsisten dengan 10 wilayah yang sama, *Cluster 2* kembali mencakup 4 wilayah, Sementara *Cluster 3* mencakup 9 wilayah.

4.3. DBSCAN

Hasil *clustering* menggunakan metode DBSCAN menunjukkan pembentukan *cluster* yang didasarkan pada kepadatan data. Tidak seperti K-Means, DBSCAN dapat menghasilkan jumlah *cluster* yang berbeda untuk tiap tahun dan memungkinkan adanya data yang diklasifikasikan sebagai *outlier* (-1). Dalam metode ini, wilayah-wilayah yang memiliki karakteristik tingkat kemiskinan yang saling berdekatan dalam ruang fitur akan dikelompokkan ke dalam suatu *cluster*, selama kepadatan titik-titik data memenuhi syarat yang ditentukan oleh *parameter eps* dan *min_samples*. Berikut hasil *clustering* setiap wilayah di Aceh menggunakan metode DBSCAN untuk tahun 2020-2024

Tabel 5. Hasil DBSCAN

Kabupaten/Kota	2020	2021	2022	2023	2024
Simeulue	0	0	0	0	0
Aceh Singkil	0	0	0	0	0
Aceh Selatan	0	1	1	0	0
Aceh Tenggara	0	1	1	0	0
Aceh Timur	0	1	2	1	1
Aceh Tengah	0	1	1	0	0
Aceh Barat	-1	-1	-1	-1	-1
Aceh Besar	0	1	2	1	1
Pidie	0	0	0	0	0
Bireuen	0	1	1	0	0
Aceh Utara	-1	-1	-1	-1	-1
Aceh Barat Daya	0	1	1	0	0

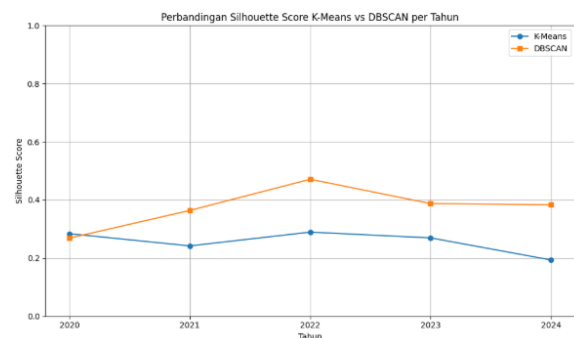
Kabupaten/Kota	2020	2021	2022	2023	2024
Gayo Lues	-1	0	0	0	0
Aceh Tamiang	0	1	2	1	1
Nagan Raya	-1	-1	-1	-1	-1
Aceh Jaya	0	1	1	0	0
Bener Meriah	-1	0	0	0	0
Pidie Jaya	0	0	0	0	0
Banda Aceh	-1	-1	-1	-1	-1
Sabang	0	1	1	0	0
Langsa	-1	1	2	1	1
Lhokseumawe	-1	-1	-1	-1	-1
Subulussalam	0	0	0	0	0

Hasil *clustering* DBSCAN tahun 2020 dengan parameter *eps* sebesar 1.0 dan *min_samples* sebanyak 5, terbentuk dua kelompok utama, yaitu *Cluster* inti (*Cluster 0*) dan sejumlah wilayah yang dikategorikan sebagai *noise* atau *outlier* (-1). *Cluster 0* merupakan *cluster* utama yang terdiri dari 15 wilayah, dan 8 kabupaten/kota yang masuk dalam kategori *noise* (-1) karena dianggap tidak memenuhi kepadatan minimum untuk membentuk suatu *cluster*. Hasil *clustering* DBSCAN tahun 2021 dengan parameter *epsilon* sebesar 1.0 dan *min_samples* sebanyak 3, terbentuk 2 *cluster* dan 1 kelompok *noise* (-1). *Cluster 0* terdiri dari 7 wilayah, *Cluster 1* terdiri dari 11 wilayah dan terdapat 5 wilayah yang dikategorikan sebagai *noise* atau *outlier*. Pada tahun 2022, algoritma DBSCAN mengelompokkan wilayah di Kabupaten/Kota Aceh menjadi 3 *Cluster* utama dan 1 kelompok *noise* atau *outlier*. *Cluster 0* mencakup 7 wilayah, *Cluster 1* terdapat 4 wilayah dan *Cluster 2* terdiri dari 4 wilayah. Selain itu, 5 lainnya terdeteksi sebagai *noise*. Hasil *clustering* DBSCAN tahun 2023 menunjukkan pola yang relatif stabil dibanding tahun sebelumnya. *Cluster 0* terdiri dari 14 wilayah, *Cluster 1* terdiri 4 wilayah Kelompok *noise* tahun ini juga tetap konsisten seperti tahun sebelumnya.

Pada tahun 2024, hasil *clustering* DBSCAN masih menunjukkan konsistensi dalam pengelompokan wilayah. *Cluster 0* merupakan *cluster* mayoritas dengan 14 wilayah, *Cluster 1* meliputi 4 wilayah kabupaten/kota dan kelompok *noise* kembali mencakup 5 wilayah yang sama seperti tahun sebelumnya.

4.4. Evaluasi Silhouette Coefficient

Evaluasi performa dari kedua metode dilakukan menggunakan *Silhouette Score*, yaitu metrik evaluasi yang mengukur seberapa baik suatu objek cocok dengan *clusternya* dibandingkan dengan *cluster* lainnya. Nilai *Silhouette Score* berada dalam rentang -1 hingga 1, di mana semakin tinggi nilainya, semakin baik hasil *clustering*. Hasil evaluasi *Silhouette Score* disajikan dalam Gambar 2. berikut



Gambar 2. Hasil evaluasi *silhouette score*

Dari diagram hasil evaluasi menggunakan *Silhouette Score* diatas, terlihat bahwa performa metode *clustering* DBSCAN lebih unggul dibandingkan K-Means. Pada tahun 2020, nilai *Silhouette Score* untuk K-Means adalah 0.2838, sedikit lebih tinggi dibandingkan DBSCAN yang memperoleh nilai 0.2686. Hal ini menunjukkan bahwa pada tahun tersebut, struktur *cluster* K-Means relatif lebih baik, meskipun selisihnya tidak terlalu besar. Namun, mulai tahun 2021 hingga 2024, DBSCAN menunjukkan peningkatan performa yang signifikan dibandingkan K-Means. Secara keseluruhan, nilai-nilai *Silhouette Score* memperlihatkan bahwa K-Means cenderung menghasilkan *cluster* yang kurang baik ketika data memiliki distribusi kompleks, atau ketika terdapat *outlier*. DBSCAN unggul dalam mengelompokkan data dengan struktur yang tidak teratur, serta mampu mengidentifikasi *noise/outlier* secara eksplisit (ditunjukkan dengan label *cluster* -1).

5. KESIMPULAN DAN SARAN

Penelitian ini bertujuan untuk mengelompokkan wilayah di Provinsi Aceh berdasarkan tiga indikator utama, yaitu persentase penduduk miskin, tingkat pengangguran, dan PDRB per kapita dengan menggunakan dua metode *clustering*, yaitu K-Means dan DBSCAN. Berdasarkan analisis yang telah dilakukan terhadap data tahun 2020 hingga 2024, K-Means *Clustering* berhasil mengelompokkan seluruh wilayah ke dalam 3 *cluster* berdasarkan jarak ke pusat *cluster*, dengan ketentuan *cluster 1* adalah kategori kemiskinan relatif tinggi, *cluster 2* adalah kategori kemiskinan rendah dan *cluster 3* adalah kategori kemiskinan sedang/menengah. Selain itu DBSCAN berhasil dalam mengidentifikasi wilayah-wilayah yang memiliki karakteristik ekstrem atau menyimpang (*outlier*), ditandai dengan keberadaan *cluster* -1. Metode ini lebih adaptif terhadap bentuk distribusi data yang tidak beraturan. Metode DBSCAN mengelompokkan wilayah di Provinsi Aceh berdasarkan kemiripan pola tingkat kemiskinannya, khususnya pada tiga variabel yang digunakan yaitu persentase penduduk miskin, tingkat pengangguran, dan PDRB per kapita. Hasil *clustering* menunjukkan pola yang cukup konsisten dari tahun ke tahun.

Berdasarkan evaluasi menggunakan *Silhouette Score*, metode DBSCAN memperoleh nilai yang lebih tinggi pada sebagian besar tahun dibandingkan K-Means, yang menunjukkan bahwa DBSCAN memiliki kualitas *clustering* yang lebih baik dalam melakukan *clustering* terhadap data kemiskinan wilayah di Aceh dari tahun 2020 hingga 2024. Meskipun pada tahun 2020 K-Means sedikit lebih baik, keunggulan tersebut tidak bertahan di tahun-tahun berikutnya. Mulai tahun 2021 hingga 2024, DBSCAN secara konsisten menunjukkan nilai *Silhouette Score* yang lebih tinggi, dengan puncak performa pada tahun 2022 sebesar 0.4711.

DAFTAR PUSTAKA

- [1] R. Y. Wulansari, N. Fadhillah, M. Huda, A. Z. Abidin, and A. E. Sujianto, "Faktor Yang Mempengaruhi Kemiskinan di Indonesia," *J. Econ. Manag. Account. Technol.*, vol. 6, no. 1, pp. 82–95, 2023, doi: 10.32500/jematech.v6i1.3928.
- [2] K. N. Sibolga Marbun and . M., "Pengaruh tenaga kerja dan PAD terhadap pertumbuhan ekonomi serta dampaknya terhadap kemiskinan di Kabupaten Ngawi," *J. Pendidik. Ekon.*, vol. 11, no. 3, pp. 320–327, 2023, doi: 10.26740/jupe.v11n3.p320-327.
- [3] Badan Pusat Statistik Provinsi Aceh, "Persentase Penduduk Miskin Aceh pada Maret 2024 sebesar 14,23 persen," Badan Pusat Statistik Provinsi Aceh. [Online]. Available: <https://aceh.bps.go.id/id/pressrelease/2024/07/01/1077/persentase-penduduk-miskin-aceh-pada-maret-2024-sebesar-14-23-persen-.html>
- [4] Muttaqien et al., "Determinasi Tingkat Kemiskinan di Kabupaten/Kota Provinsi Aceh," *J. EMT KITA*, vol. 7, no. 3, pp. 763–770, 2023, doi: 10.35870/emt.v7i3.1250.
- [5] M. Balqis, N. Q. Sinaga, and N. Hanum, "Pengaruh Pendidikan dan Upah Minimum Terhadap Kemiskinan di Provinsi Aceh," 2025.
- [6] W. Setya and A. Nugraha, "Clustering Pemetaan Tingkat Kemiskinan di Provinsi Jawa Barat Menggunakan Algoritma K-Means," *J. Ilm. Wahana Pendidikan*, Januari, vol. 2023, no. 2, pp. 234–244, 2023, [Online]. Available: <https://doi.org/10.5281/zenodo.7567622>.
- [7] J. Ramadhani, Y. Safitra Anugraha, A. Fauzan, and L. Efrizoni, "Perbandingan Algoritma K-Means Clustering Dan K-Medoids Dalam Mengelompokkan Tingkat Kemiskinan Di Provinsi Riau," *J. Jar. Sist. Inf. Robot.*, vol. 8, no. 1, pp. 114–125, 2024, [Online]. Available: <https://ojsamik.amikmitragama.ac.id/index.php/js/article/view/393>
- [8] B. Anthony et al., "Zeta – Math Journal," vol. 9, no. 2023, pp. 31–40, 2024.
- [9] S. P. Dewi, N. Nurwati, and E. Rahayu, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor," *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.
- [10] S. Suliman, "Implementasi Data Mining Terhadap Prestasi Belajar Mahasiswa Berdasarkan Pergaulan dan Sosial Ekonomi Dengan Algoritma K-Means Clustering," *Simkom*, vol. 6, no. 1, pp. 1–11, 2021, doi: 10.51717/simkom.v6i1.48.
- [11] M. Ula, R. Maulana, and V. Ilhadi, "Penerapan KNN Penentuan Pelanggan Baru PDAM dan Clustering K-Means Berdasarkan Wilayah," vol. 04, no. 01, pp. 1–6, 2023, [Online]. Available: <https://ejournalunsam.id/index.php/jicom/>
- [12] R. Meiyanti, M. M. Munauwar, R. Fitria, and H. Al Kautsar, "Implementasi Algoritma K-Medoid pada Clustering Sayuran Unggulan di Kabupaten Aceh Utara," vol. 19, no. x, pp. 327–337, 2024.
- [13] S. N. Mayasari and J. Nugraha, "Implementasi K-Means Cluster Analysis untuk Mengelompokkan Kabupaten/Kota Berdasarkan Data Kemiskinan di Provinsi Jawa Tengah Tahun 2022," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 317–329, 2023, doi: 10.24002/konstelasi.v3i2.7200.
- [14] F. O. Dayera, Musa Bundaris Palungan, "G-Tech : Jurnal Teknologi Terapan," *G-Tech J. Teknol. Terap.*, vol. 8, no. 1, pp. 186–195, 2024, [Online]. Available: <https://ejournal.uniramalang.ac.id/index.php/g-tech/article/view/1823/1229>
- [15] B. Anthony et al., "Zeta – Math Journal," vol. 9, no. 2023, pp. 31–40, 2024.