JISTech (Journal of Islamic Science and Technology)

ISSN: 2528-5718

Vol. 11, No. 1, 2026. Page: 105-117

http://jurnal.uinsu.ac.id/index.php/jistech

Penerapan *Principal Component Analysis* dan *Cluster Analysis* untuk Segmentasi Kabupaten/Kota di Sumatera Utara Berdasarkan Indikator Pembangunan Manusia

Elma Dwi Ariana Aprilia Zam[1*], Hani Maulida Hasibuan[2], Aida Febriana Tanjung[3], Ahmad Syahrial Pohan[4], Rina Filia Sari[5]

[1,2,3,4,5] *Universitas Islam Negeri Sumatera Utara, Medan, Indonesia*

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Sumatera Utara berdasarkan indikator pembangunan manusia menggunakan kombinasi metode *Principal component analysis* (PCA) dan *K-means Clustering*. Data yang digunakan merupakan data indikator pembangunan manusia pada 33 kabupaten/kota di Provinsi Sumatera Utara. Variabel yang digunakan adalah harapan lama sekolah (x_1), rata-rata lama sekolah (x_2), angka harapan hidup (x_3), tingkat pengangguran terbuka (x_4), jumlah penduduk miskin (x_5), dan Pengeluaran perkapita (x_6). Analisis diawali dengan reduksi dimensi menggunakan PCA untuk mengatasi korelasi antar variabel dan memperoleh faktor utama yang mewakili karakteristik data. Hasil pengujian menunjukkan bahwa satu komponen utama dengan nilai eigen sebesar 2.848 mampu menjelaskan 56.963% keragaman data. Faktor tersebut dibentuk oleh variabel harapan lama sekolah, rata-rata lama sekolah, angka harapan hidup, tingkat pengangguran terbuka, dan pengeluaran per kapita. Selanjutnya, skor faktor hasil PCA digunakan sebagai input pada metode *K-means Clustering*. Berdasarkan metode elbow, jumlah *cluster* optimal yang diperoleh adalah empat *cluster*. Hasil pengelompokan menunjukkan bahwa *Cluster* 1 termasuk kategori rendah dengan 4 kabupaten/kota, *Cluster* 3 kategori sedang dengan 15 kabupaten/kota, *Cluster* 2 kategori tinggi dengan 2 kabupaten/kota, dan *Cluster* 4 kategori sangat tinggi dengan 12 kabupaten/kota. Hasil penelitian menunjukkan adanya perbedaan karakteristik indikator pembangunan manusia antar kelompok kabupaten/kota di Provinsi Sumatera Utara sehingga dapat menjadi dasar dalam penyusunan kebijakan pembangunan yang lebih terarah.

ABSTRACT

This study aimed to classify regencies and municipalities in North Sumatra Province based on human development indicators using a combination of Principal component analysis (PCA) and K-Means Clustering. The data consisted of human development indicators from 33 regencies and municipalities in North Sumatra Province. PCA was first employed to reduce data dimensionality and address correlations among variables by extracting the most representative factor. The results indicated that one principal component with an eigenvalue of 2.848 explained 56.963% of the total variance. This component was formed by the variables of Expected Years of Schooling, Mean Years of Schooling, Life Expectancy, Open Unemployment Rate, and Expenditure per Capita. The factor scores obtained from PCA were subsequently used as input for the K-Means clustering method. Based on the elbow method, the optimal number of clusters was determined to be four. The clustering results showed that Cluster 1 was categorized as low with 4 regencies/municipalities, Cluster 3 as moderate with 15 regencies/municipalities, Cluster 2 as high with 2 regencies/municipalities, and Cluster 4 as very high with 12 regencies/municipalities. These findings indicate the existence of distinct human development characteristics among regencies and municipalities in North Sumatra Province, which may serve as a basis for formulating more targeted development policies.

Kata Kunci: *Principal Component Analysis*; *K-Means Clustering*; Indikator Pembangunan Manusia; Reduksi Dimensi; Pengelompokkan
Email: * elma0706253003@uinsu.ac.id

DOI: http://dx.doi.org/10.30829/jistech.v11i1.30953

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Pendahuluan

Pembangunan manusia merupakan salah satu indikator penting dalam mengukur keberhasilan pembangunan suatu wilayah. Konsep pembangunan manusia tidak hanya berorientasi pada pertumbuhan ekonomi, tetapi juga

menekankan peningkatan kualitas hidup masyarakat melalui akses terhadap pendidikan, kesempatan kerja, dan kesejahteraan ekonomi [1]. Keberhasilan pembangunan manusia di suatu daerah dapat mencerminkan kemampuan masyarakat dalam memperoleh kehidupan yang layak dan produktif [2].

Kondisi pembangunan manusia dapat ditinjau melalui berbagai indikator pembangunan manusia [3]. Angka Harapan Hidup (AHH) merupakan indikator kesehatan yang menggambarkan rata-rata jumlah tahun hidup yang diharapkan dapat dijalani oleh penduduk sejak lahir dengan mempertimbangkan pola mortalitas yang berlaku pada suatu periode tertentu. Sementara itu, Rata-rata Lama Sekolah (RLS) dan Harapan Lama Sekolah (HLS) merupakan indikator pendidikan yang mencerminkan capaian pendidikan penduduk serta peluang penduduk usia sekolah untuk memperoleh pendidikan di masa mendatang. Tingkat Pengangguran Terbuka (TPT) mencerminkan kemampuan pasar kerja dalam menyerap tenaga kerja yang tersedia. Pengeluaran per kapita menunjukkan tingkat kesejahteraan masyarakat, sedangkan jumlah penduduk miskin menggambarkan tingkat kerentanan sosial ekonomi suatu wilayah. Keenam indikator tersebut mampu memberikan gambaran yang lebih komprehensif mengenai kualitas pembangunan manusia karena mencakup aspek kesehatan, pendidikan, ketenagakerjaan, kesejahteraan, dan kemiskinan yang saling memengaruhi dalam proses pembangunan suatu daerah.

Pengelompokan kabupaten/kota berdasarkan kemiripan karakteristik indikator pembangunan manusia dapat menjadi salah satu pendekatan untuk mendukung proses perencanaan, pelaksanaan, dan evaluasi pembangunan di Provinsi Sumatera Utara. Melalui pengelompokan tersebut, pemerintah daerah dapat mengidentifikasi wilayah yang memiliki karakteristik serupa sehingga kebijakan yang dirumuskan dapat lebih efektif dan sesuai dengan kebutuhan masing-masing kelompok wilayah. Selain itu, hasil pengelompokan dapat digunakan sebagai dasar dalam upaya meningkatkan kualitas pembangunan manusia sekaligus mengurangi kesenjangan pembangunan antar kabupaten/kota di Provinsi Sumatera Utara.

Beberapa penelitian terdahulu telah menerapkan analisis klaster pada indikator pembangunan manusia. Penelitian [4] mengelompokkan kabupaten/kota di Indonesia berdasarkan indikator IPM menggunakan metode K-means dan Fuzzy C-Means, serta menunjukkan adanya perbedaan karakteristik pembangunan manusia antar kelompok wilayah. Penelitian [5] menggunakan metode K-means untuk mengklasifikasikan kabupaten/kota di Indonesia berdasarkan dimensi IPM dan memperoleh empat kelompok wilayah dengan tingkat pembangunan yang berbeda. Selain itu, penelitian yang dilakukan oleh [6] mengombinasikan PCA dan K-means *Clustering* untuk mengelompokkan negara berdasarkan indikator pembangunan global, yang menunjukkan bahwa PCA mampu meningkatkan efektivitas proses pengelompokan dengan mereduksi dimensi data terlebih dahulu. Keunggulan PCA terletak pada kemampuannya dalam menyederhanakan kompleksitas data tanpa memerlukan keberadaan variabel dependen tertentu, sehingga metode ini sangat sesuai untuk digunakan dalam analisis eksploratif [7].

Dalam penelitian ini, digunakan pendekatan *Principal component analysis* (PCA) dan *Cluster* Analysis untuk melakukan segmentasi kabupaten/kota di Provinsi Sumatera Utara berdasarkan indikator Tingkat Pengangguran Terbuka (TPT), Angka Harapan Hidup (AHH), Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), pengeluaran per kapita, dan jumlah penduduk miskin, lalu dilakukan klaster menggunakan metode k-means *clustering*. PCA digunakan untuk mereduksi dimensi data dan mengatasi korelasi antarvariabel sehingga diperoleh sejumlah komponen utama yang mampu merepresentasikan sebagian besar keragaman data [8].

Clustering merupakan teknik analisis data yang digunakan untuk mengelompokkan objek berdasarkan kemiripannya [9]. Analisis klaster berujuan mengelompokkan objek ke dalam beberapa kelompok berdasarkan kemiripan karakteristik, setiap objek akan dimasukkan ke kelompok yang jaraknya paling dekat, sehingga antar kelompok memiliki perbedaan yang mencolok [10]. Analisis klaster menggunakan metode k-means diterapkan untuk mengelompokkan kabupaten/kota Provinsi Sumatera Utara yang memiliki karakteristik indikator pembangunan manusia yang serupa [11]. K-means *Clustering* adalah metode analisis *cluster* non-hierarki yang bertujuan mengelompokkan objek ke dalam beberapa kelompok sehingga objek dalam *cluster* yang sama memiliki karakteristik yang lebih mirip dibandingkan dengan objek pada *cluster* lainnya [12]. Di antara berbagai algoritma yang tersedia, K-means menjadi salah satu metode yang populer karena mampu memproses data skala besar secara cepat dan efisien [13]. Jumlah klaster optimal ditentukan menggunakan metode Elbow sehingga hasil segmentasi yang diperoleh dapat menggambarkan struktur data secara lebih representatif. Hasil penelitian ini diharapkan dapat menjadi bahan pertimbangan bagi pemerintah daerah dalam menyusun kebijakan pembangunan yang lebih terarah, efektif, dan sesuai dengan karakteristik masing-masing kelompok kabupaten/kota di Provinsi Sumatera Utara.

Metodologi Penelitian

Penelitian ini merupakan penelitian kuantitatif. Data yang digunakan adalah data sekunder tahun 2023 yang diperoleh dari website Badan Pusat Ststistik Provinsi Sumatera Utara. Data tersebut digunakan untuk menggambarkan karakteristik pembangunan manusia pada 33 kabupaten/kota di Provinsi Sumatera Utara tahun 2023. Variable yang digunakan yaitu Harapan Lama Sekolah (x_1), Rata-Rata Lama Sekolah (x_2), Angka Harapan Hidup (x_3), Tingkat Pengangguran Terbuka (x_4), Persentase Penduduk Miskin (x_5), dan Pengeluaran perkapita (x_6). Metode yang diguanakan adalah *Principal component analysis* (PCA) dan *K-means Clustering*.

Principal component analysis (PCA) merupakan teknik statistika multivariat yang digunakan untuk mereduksi dimensi data dengan mengubah sejumlah variabel asal menjadi beberapa komponen utama yang saling bebas korelasi. Metode ini bekerja melalui transformasi variabel-variabel yang memiliki hubungan korelasi menjadi sekumpulan variabel baru yang dikenal sebagai komponen utama (*principal components*). Pembentukan komponen

utama didasarkan pada nilai eigen yang diperoleh dari matriks korelasi, di mana komponen dengan nilai eigen lebih besar atau sama dengan satu dipertahankan karena dianggap mampu merepresentasikan informasi penting dalam data [14]. Adapun langkah-langkah PCA yaitu sebagai berikut:

1. Pengumpulan Data
2. Standarisasi Data
   Standarisasi data dilakukan sebelum menghitung PCA karena satuan pengukuran setiap variabel berbeda [15] menggunakan rumus:
   $$Z_{ji} = \frac{(x_{ji} - \bar{x}_j)}{s_k}$$
3. Membentuk matriks korelasi atau kovarians menggunakan yang ditunjukkan seperti dibawah ini.
   $$Cov(X) = \frac{1}{n-1} X^T X$$
   X adalah matriks data setelah dikurangi oleh rata-rata n jumlah observasi.
4. Hitung Nilai Eigen dan Vektor Eigen
   Selanjutnya dihitung *eigenvalue* (λ) dan *eigenvector* (v) dari matriks yang sudah dihitung kovarians atau korelasi dengan menyelesaikan:
   $$Cov(X) \cdot v = \lambda \cdot v$$
   $$det(Cov(X) - \lambda I) = 0$$
   Nilai *Eigen* menunjukkan seberapa besar varians yang dijelaskan setiap komponen. Sementara *vektor eigen* menunjukkan kombinasi dari variable dari komponen.
5. Memilih Komponen Utama
   Nilai *eigen* diurutkan dari yang terbesar ke yang terkecil. Lalu nilai komponen utama (*Principal Components*) dengan nilai *eigen* ≥ 1 dalam kaidah *Kaiser* menjelaskan bahwa hanya 70-90% varians kumulatif.
6. Membuat scree plot.
7. Membentuk factor loading matriks untuk interpretasi.
8. Penentuan jumlah komponen utama.

Selanjutnya, hasil reduksi dimensi yang diperoleh dari PCA digunakan sebagai dasar dalam proses pengelompokan data (*clustering*) [16]. Salah satu masalah utama dalam *cluster* adalah penentuan jumlah klister optimal (k) secara otomatis tanpa perlu inisialisasi manual. Untuk mengatasi hal ini, beberapa metode telah dikembangkan, salah satunya adalah metode *Elbow*, misalnya memanfaatkan nilai *Within Cluster Sum of Square* (WCSS) untuk mendeteksi nilai titik siku sebagai indikator jumlah klister terbaik. Adapun tahapan analisis nya yaitu [17]:

1. Menetukan jumlah klaster (k)
2. Inisialisasi titik pusat awal (*Centroid*)
3. Menghitung Jarak tiap titik ke setiap *centroid* menggunakan jarak *euclid* dengan rumus:
   $$d(x_i, c_j) = \sqrt{\sum_{m=1}^{n} (x_{im} - c_{jm})^2}$$
   Dimana x_i data ke i, c_j centroid klaster ke-j, n jumlah dimensi fitur
4. Ulangi langkah 3-5 (iterasi), mengelompokkan data dan menghitung *centroid* baru sampai:
   a. Poisis *centroid* sudah konsisten atau tidak berubah.
   b. Jumlah perulangan mencapai maksimum.
   c. Anggota klaster sudah tidak mengalami perubahan.
5. Evaluasi hasil dengan WCSS menggunakan rumus sebagai berikut.
   $$WCSS = \sum_{j=1}^{k} \sum_{x_i \epsilon C_j} \|x_i - c_j\|^2$$

Semakin kecil nilai WCSS menandakan semakin baik hasil klasterisasi karena titik-titik lebih dekat ke nilai pusatnya.

Hasil dan Pembahasan

Data yang digunakan diperoleh dari website resmi Badan Pusat Statistik (BPS) Sumatera Utara https://sumut.bps.go.id. Data yang digunakan adalah periode tahun 2023. Berikut data indikator pembangunan manusia yang disajikan dalam Tabel 1.

Tabel 1. Data Penelitian

Kab/Kota	HLS	RLS	AHH	TPT	Penduduk Miskin	Pengeluaran perkapita
Nias	13.3	6.14	71.74	2.31	15,10	7301
Mandailing Natal	13.86	8.84	71.72	7.45	8,86	10251
Tapanuli Selatan	13.58	9.51	71.61	3.49	7,01	11829
Tapanuli Tengah	13.49	8.87	71.76	7.81	11,50	10690
Tapanuli Utara	13.73	10.09	74.1	1.03	8,54	12115
Toba	13.59	10.59	74.22	1.3	8,04	12676
Labuhan Batu	13.25	9.49	72.88	5.99	7,99	11670
Asahan	12.64	8.83	73.39	6.12	8,21	11795
Simalungun	12.82	9.72	74.08	5.35	7,87	11746
Dairi	13.32	9.88	74.13	1.23	7,47	10969
Karo	13.25	10.03	74.16	2.63	7,98	12779
Deli Serdang	13.39	10.28	73.65	8.62	3,44	12890
Langkat	13.27	8.73	74.14	6.33	9,23	11632
Nias Selatan	12.78	6.48	71.61	3.48	16,39	7299
Humbang Hasundutan	13.32	10.01	74.07	0.84	8,69	8476
Pakpak Bharat	13.9	9.61	72.61	0.45	7,54	8764
Samosir	13.51	9.47	74.1	1.03	11,66	9158
Serdang Bedagai	12.64	8.85	73.11	4.97	7,44	11695
Batu Bara	13.11	8.5	72.63	5.88	11,38	10933
Padang Lawas Utara	13.53	9.55	71.57	4.42	8,79	10615
Padang Lawas	13.7	9.43	71.52	5.75	7,89	9395
Labuhanbatu Selatan	13.42	8.93	72.46	3.43	8,06	11950
Labuanbatu Utara	13.57	8.87	74.06	4.84	9,08	12429
Nias Utara	13.36	6.85	72.12	2.57	21,79	6788
Nias Barat	12.98	7.07	72.3	0.8	22,81	6382
Sibolga	13.42	10.44	74.02	6.79	11,42	12285
Tanjungbalai	13.14	9.68	74.01	4.47	12,21	11753
Pematangsiantar	14.6	11.58	74.75	8.62	7,24	12984
Tebing Tinggi	13.12	10.86	74.07	6.24	9,49	13385
Medan	14.78	11.62	74.76	8.67	8,00	15674
Binjai	14.17	11.19	74.18	6.1	4,79	11567
Padangsidimpuan	14.59	11.12	73.54	7.57	6,85	11552
Gunungsitoli	13.78	8.65	74.03	3.67	14,78	8635

Sumber: Website BPS Sumut

Ruang lingkup dimensi variabel yang digunakan yaitu:

1. Dimensi Pendidikan: Harapaan Lama Sekolah (x_1), Rata-Rata Lama Sekolah (x_2)
2. Dimensi Kesehatan: Angka Harapan Hidup (x_3)
3. Dimensi Ketenagakerjaan: Tingkat Pengangguran Terbuka (x_4)
4. Dimensi Kesejahteraan Sosial: Persentase Penduduk Miskin (x_5)
5. Dimensi Ekonomi: Pengeluaran perkapita (x_6)

Selanjutnya akan dilakukan analisis deskriptif yaitu sebagai berikut.

Tabel 2. Analisis Deskriptif Data Penelitian

Variabel	Min	Max	Mean	Std. Deviasi
X_1	12.64	14.78	13.4821	0.51356
X_2	6.14	11.62	9.3867	1.34188
X_3	71.52	74.76	73.2455	1.07088
X_4	0.45	8.67	4.5530	2.55136
X_5	3.44	22.81	9.9255	4.19192
X_6	6382	15674	10910.97	2100.480

Sumber : Hasil Olah Data (Output SPSS)

Berdasarkan statistik deskriptif, $x_1$memiliki rata-rata 13,4821 dengan simpangan baku 0,51356, sehingga menunjukkan variasi data yang rendah. Variabel $x_2$dan x_3 juga memiliki penyebaran yang relatif kecil, masing-masing dengan rata-rata 9,3867 dan 73,2455, serta simpangan baku 1,34188 dan 1,07088, yang mengindikasikan data relatif homogen. Sementara itu, x_4 dan x_5 memiliki keragaman yang lebih tinggi, ditunjukkan oleh simpangan baku 2,55136 dan 4,19192. Variabel x_6 memiliki rata-rata terbesar, yaitu 10.910,97, dengan simpangan baku 2.100,480, sehingga merupakan variabel dengan tingkat penyebaran data paling tinggi. Secara keseluruhan, hasil

statistik deskriptif menunjukkan adanya perbedaan tingkat variasi antarvariabel yang perlu diperhatikan dalam analisis selanjutnya.

Standarisasi Data

Sebelum dilakukan *Principal component analysis* (PCA), data terlebih dahulu melalui tahap preprocessing. Pada tahap ini, variabel Tingkat Pengangguran Terbuka (TPT) dan persentase penduduk miskin ditransformasikan dengan mengalikan setiap nilai pengamatan dengan -1. Transformasi tersebut dilakukan karena kedua variabel merupakan indikator yang bersifat negatif, di mana nilai yang lebih tinggi menunjukkan kondisi pembangunan yang kurang baik. Dengan demikian, setelah transformasi seluruh variabel memiliki arah interpretasi yang sama, yaitu nilai yang lebih besar merepresentasikan kondisi pembangunan yang lebih baik. Selanjutnya, seluruh variabel distandarisasi menggunakan Z_{score} agar memiliki skala pengukuran yang seragam dan menghilangkan pengaruh perbedaan satuan antarvariabel. Data hasil standarisasi kemudian digunakan sebagai masukan dalam proses *Principal component analysis* (PCA).

Tabel 3. Data Standarisasi

x_1	x_2	x_3	x_4	x_5	x_6
-0.35462	-2.41949	-1.40582	0.87915	-1.23441	-1.71864
0.73580	-0.40739	-1.42449	-1.13546	0.25417	-0.31420
0.19059	0.09191	-1.52721	0.41665	0.69549	0.43706
0.01534	-0.38503	-1.38714	-1.27656	-0.37561	-0.10520
0.48266	0.52414	0.79799	1.38085	0.33051	0.57322
⋮	⋮	⋮	⋮	⋮	⋮
-0.70511	1.09796	0.76997	-.66120	0.10388	1.17784
2.52720	1.66433	1.41431	-1.61364	0.45932	2.26759
1.33942	1.34389	0.87269	-0.60633	1.22508	0.31232
2.15723	1.29172	0.27505	-1.18250	0.73366	0.30518
0.58002	-0.54898	0.73262	0.34610	-1.15807	-1.08355

Sumber : Hasil Olah Data (Output SPSS)

Menghitung Matriks Korelasi

Matriks korelasi memiliki nilai yang sama dengan matriks varians-kovarians dari data yang sudah distandarisasi.

Tabel 4. Matriks Korelasi Data Standarisasi

x_1	x_2	x_3	x_4	x_5	x_6
1	0.573	0.245	-0.304	0.307	0.297
0.573	1	0.653	-0.370	0.744	0.772
0.245	0.653	1	-0.093	0.348	0.569
-0.304	-0.370	-0.093	1	-0.369	-0.550
0.307	0.744	0.348	-0.484	1	0.721
0.297	0.772	0.569	-0.550	0.721	1

Sumber : Hasil Olah Data (Output SPSS)

Berdasarkan Tabel 4, matriks korelasi menunjukkan bahwa hubungan antarvariabel bersifat positif maupun negatif dengan kekuatan yang bervariasi. Korelasi positif terkuat diperoleh antara x_2 dan x_6 (r = 0,772), diikuti oleh x_2 dan x_5 (r = 0,744) serta $x_2$dan x_6 (r = 0,721). Nilai korelasi tersebut mengindikasikan hubungan linier yang kuat, sehingga peningkatan pada salah satu variabel cenderung diikuti oleh peningkatan pada variabel lainnya. Selain itu, hubungan antara x_2 dan x_3 (r = 0,653) tergolong cukup kuat, sedangkan hubungan yang melibatkan x_1 umumnya berada pada kategori lemah hingga sedang.

Dimana $x_4$memiliki korelasi negatif terhadap seluruh variabel lainnya, dengan nilai berkisar antara -0,093 hingga -0,550. Korelasi negatif terkuat terdapat antara X_4 dan X_6 (r = -0,550), diikuti oleh $x_4$dan x_5 (r = -0,484) serta x_4 dan x_2 (r = -0,370). Kondisi ini menunjukkan bahwa peningkatan nilai X_4 cenderung diikuti oleh penurunan nilai pada variabel-variabel tersebut.

Secara umum, pola korelasi memperlihatkan bahwa beberapa variabel memiliki keterkaitan yang cukup erat, sedangkan variabel lainnya memberikan informasi yang relatif berbeda. Keberadaan korelasi yang cukup kuat ($|r| \geq 0,70$) mengindikasikan adanya redundansi informasi antarbeberapa variabel. Oleh karena itu, data memenuhi karakteristik yang sesuai untuk diterapkan *Principal component analysis* (PCA) atau analisis biplot, karena kedua metode tersebut efektif dalam mereduksi dimensi data sekaligus mempertahankan sebagian besar keragaman informasi yang terkandung dalam variabel-variabel penelitian.

Uji KMO & Barlett's

Data diuji kelayakannya untuk memastikan adanya korelasi antarvariabel, Uji *Kaiser-Meyer-Olkin* (KMO) digunakan untuk menilai kecukupan sampel dan kecocokan data untuk dianalisis PCA [18], sedangkan Bartlett's Test of Sphericity digunakan untuk menguji apakah matriks korelasi berbeda dari matriks identitas [19]. PCA layak

digunakan jika nilai KMO > 0,50 dan Nilai Barlett's Test p < 0,05. Hasil uji KMO dan Barlett disajikan pada **Tabel 5**.

Tabel 5. Uji KMO & Barlett's

Indikator	Nilai
Kaiser-Meyer-Olkin Measure	0.669
Bartlett's Test of Sphericity	106.055
df	15
Sig.	0.000

Sumber : Hasil Olah Data (Output SPSS)

Hasil perhitungan dengan SPSS dihasilkan dari nilai *Bartlett's Test of Sphericity* sebesar 90.791 dengan taraf signifikan < α (0.05) dan nilai KMO diatas 0.5, maka kumpulan variabel tersebut dapat diproses lebih lanjut.

Nilai Measure of Sampling Adequancy (MSA)

Selanjutnya dilakukan evaluasi terhadap masing-masing variabel untuk mengetahui tingkat kelayakannya dalam pembentukan komponen utama. Setelah data memenuhi persyaratan berdasarkan uji Bartlett dan Kaiser-Meyer-Olkin (KMO), pengujian dilanjutkan dengan Measure of Sampling Adequacy (MSA). Nilai MSA digunakan sebagai dasar dalam menentukan variabel yang tetap dipertahankan atau dikeluarkan dari analisis. Adapun hasil pengujian MSA menggunakan software SPSS ditampilkan pada **Tabel 6** berikut.

Tabel 6. Nilai Uji MSA

Variabel	Nilai MSA
X_1	0.557
X_2	0676
X_3	0.611
X_4	0.621
X_5	0.711
X_6	0.738

Sumber: Hasil Olah Data (Output SPSS)

Berdasarkan **Tabel 6**, dapat diketahui bahwa tidak terdapat nilai MSA < 0.5. Sehingga smeua variabel memenuhi syarat MSA, sehingga semua variabel layak diikutsertakan ke analisis berikutnya.

Komunalitas

Nilai komunalitas yang tinggi menunjukkan bahwa variabel tersebut berkontribusi dengan baik dalam pembentukan struktur PCA.

Tabel 7. Nilai Komunalitas

Variabel	Komunalitas	Initial
X_1	0.362	1.000
X_2	0.901	1.000
X_3	0.818	1.000
X_4	0.827	1.000
X_5	0.657	1.000
X_6	0.802	1.000

Sumber: Hasil Olah Data (output SPSS)

Berdasarkan tabel 7 diatas, terlihat kolom initial memiliki nilai 1 untuk setiap variabel. Angka ini terlihat didalam diagonal matriks korelasi. Sedangkan pada kolom komunalitas menunjukkan seberapa besar faktor yang terbentuk dapat menerangkan varian suatu variabel. Nilai communalities tertinggi adalah variabel x_2 sebesar 0.901 artinya faktor rata-rata lama sekolah dapat menjelaskan 90.1% varians faktor yang terbentuk. Sebaliknya nilai communalities yang terendah adalah variabel x_1 sebesar 0,362 artinya harapan lama sekolah dapat menjelaskan 36.2% varians faktor yang terbentuk. Begitu seterusnya untuk variabel-variabel yang lain.

Total Varians

Tabel 8. Analisis Eigenvalue dan Proporsi Varians

Komponen	Eigenvalue	Proporsi_Varian	Kumulatif_Varian
PC1	3.418	56.964	56.964
PC2	0.950	15.829	72.793

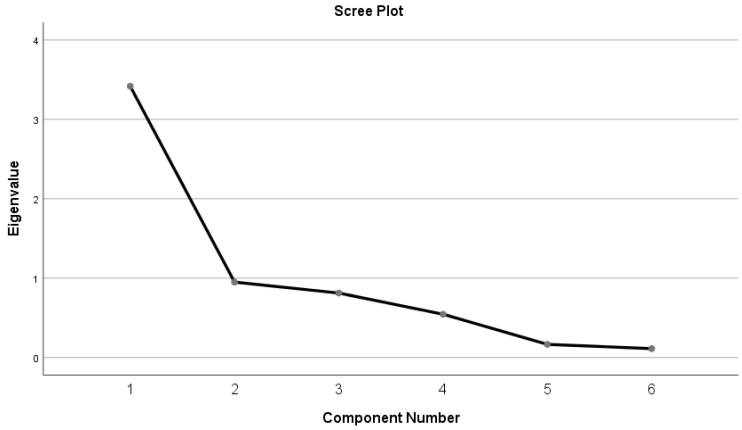
PC3	0.811	13.523	86.316
PC4	0.545	9.079	95.395
PC5	0.165	2.749	98.144
PC6	0.111	1.856	100.000

Sumber: Hasil Olah Data (output SPSS)

Berdasarkan Tabel 8 hasil *Principal component analysis* (PCA), diperoleh enam komponen utama dengan nilai eigenvalue dan proporsi keragaman yang berbeda-beda. Komponen utama pertama (PC1) memiliki eigenvalue sebesar 3,418 dan mampu menjelaskan 56,964% dari total keragaman data. Komponen utama kedua (PC2) memiliki eigenvalue sebesar 0,950 dengan kontribusi keragaman sebesar 15,829%. Secara kumulatif, PC1 dan PC2 mampu menjelaskan 72,793% dari total variasi data, sehingga kedua komponen tersebut telah merepresentasikan sebagian besar informasi yang terkandung dalam enam variabel penelitian.

Meskipun berdasarkan kriteria Kaiser hanya PC1 yang memiliki eigenvalue lebih besar dari 1, pada penelitian ini dua komponen utama dipertahankan karena secara kumulatif telah mampu menjelaskan 72,793% dari total keragaman data. Persentase tersebut telah melebihi batas yang umum digunakan dalam analisis multivariat, yaitu sekitar 70%, sehingga dua komponen utama dianggap telah cukup merepresentasikan informasi dari variabel asli dengan kehilangan informasi yang relatif kecil. Selanjutnya, kedua skor komponen utama tersebut digunakan sebagai variabel masukan pada analisis *K-Means Clustering*.

.



Gambar 1. Scree Plot

Gambar Scree Plot menunjukkan bahwa terjadi penurunan nilai eigenvalue yang sangat tajam dari komponen pertama (PC1) ke komponen kedua (PC2), kemudian setelah komponen kedua penurunan nilai eigenvalue cenderung melandai. Pola tersebut mengindikasikan bahwa sebagian besar keragaman data telah dijelaskan oleh komponen pertama, sedangkan komponen-komponen berikutnya memberikan tambahan informasi yang relatif lebih kecil.

Hal ini sejalan dengan hasil proporsi keragaman kumulatif, di mana PC1 dan PC2 mampu menjelaskan sebesar 72,793% dari total keragaman data, sehingga kedua komponen tersebut dinilai telah cukup merepresentasikan informasi yang terkandung dalam enam variabel penelitian.

Komponen Matriks

Tabel 9. Nilai Komponen Matriks

Variabel	Komponen	
	1	**2**
X_1	0.573	-0.185
X_2	0.936	0.159
X_3	0.660	0.619
X_4	-0.568	0.711
X_5	0.809	-0.039
X_6	0.895	-0.018

Sumber: Hasil Olah Data (output SPSS)

Hasil analisis menunjukkan bahwa Tabel 9, Komponen 1 memiliki loading terbesar pada x_2 (0,936), x_6 (0,895), dan x_5 (0,809), diikuti x_3 (0,660) dan x_1 (0,573), sehingga variabel-variabel tersebut merupakan penyusun utama komponen pertama. Pada variabel x_4 memiliki loading negatif (-0,568), yang menunjukkan arah hubungan yang berlawanan terhadap Komponen 1. Pada Komponen 2, loading terbesar terdapat pada x_4 (0,711) dan x_3 (0,619),

sedangkan variabel lainnya memiliki loading yang relatif kecil sehingga kontribusinya terhadap komponen ini lebih rendah. Secara keseluruhan, hasil matriks komponen menunjukkan bahwa Komponen 1 merepresentasikan variasi utama data yang didominasi oleh x_2 , x_5 , dan x_6, sedangkan Komponen 2 terutama dibentuk oleh x_4 dan x_3, sehingga kedua komponen mampu merangkum informasi utama dari seluruh variabel penelitian.

Data pada penelitian ini diproses mengggunakan PCA kemudian disimpan dalam 2 komponen utama. Data hasil PCA ini yang kemudian akan digunakan pada proses *clustering*. Data hasil proses PCA terdapat pada **Tabel 10** berikut.

Tabel 10. Data Hasil Proses PCA

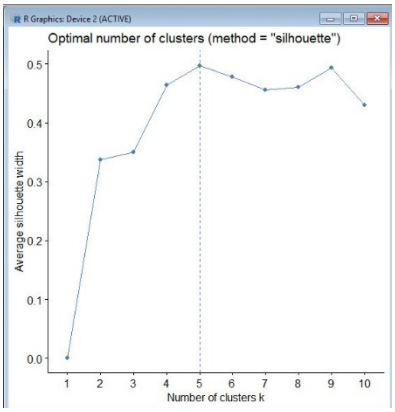
Provinsi	Faktor 1	Faktor 2
Nias	-1,88178	-0,51131
Mandailing Natal	-0,09677	-1,99454
Tapanuli Selatan	-0,02769	-0,74217
Tapanuli Tengah	-0,27512	-1,90924
Tapanuli Utara	0,37750	1,52260
Toba	0,57127	1,62216
Labuhan Batu	0,17703	-0,56859
Asahan	-0,05308	-0,14589
Simalungun	0,27459	0,53965
Dairi	0,13687	1,61140
Karo	0,43809	1,25333
Deli Serdang	1,10310	-0,88174
Langkat	0,20294	0,00850
Nias Selatan	-2,00245	-0,68073
Humbang Hasundutan	-0,25262	1,73970
Pakpak Bharat	-0,33269	0,68133
Samosir	-0,36580	1,58585
Serdang Bedagai	-0,14329	0,01680
Batu Bara	-0,40630	-0,71887
Padang Lawas Utara	-0,23447	-0,98693
Padang Lawas	-0,22723	-1,48511
Labuhanbatu Selatan	-0,09316	-0,20965
Labuanbatu Utara	0,32582	0,29175
Nias Utara	-2,07387	-0,21984
Nias Barat	-2,34398	0,59443
Sibolga	0,56679	-0,02707
Tanjungbalai	0,05658	0,67135
Pematangsiantar	1,75838	-0,47318
Tebing Tinggi	0,77406	0,30232
Medan	2,12284	-0,56156
Binjai	1,23347	0,02220
Padangsidimpuan	1,21827	-0,94632
Gunungsitoli	-0,52731	0,59936

Sumber: Hasil Olah Data (output SPSS)

Clustering K-means

Proses *clustering* dilakukan terhadap data hasil proses PCA yang menghasilkan data dengan 2 komponen yaitu PC1 dan PC2. *Clustering* dianalisis menggunakan metode non-hirarki yaitu algoritma *k-means*. Penentuan jumlah klaster pada algoritma *k-means* menggunakan metode *elbow* [20]. Berikut langkah-langkah untuk *clustering*:

1. Metode *Silhouette*



Gambar 2. Visualisasi k Optimal Metode *Silhouette*

Berdasarkan Gambar 2 dengan menggunakan metode *silhouette*, lima *cluster* adalah jumlah klaster yang ideal.

1. Analisis Klaster dengan *K-means*

Pertama dilakukan proses iterasi, yaitu proses pengacakan yang mencoba-coba mengubah atau mengacak *cluster* yang sebelumnya (initial) sehingga menjadi lebih tepat dalam pengelompokan 33 kabupaten/kota.

Tabel 11. Initial *Cluster* Centers

	Cluster				
	1	**2**	**3**	**4**	**5**
Factor 1	-2.34398	-0.09677	-0.14329	-0.25262	2.12284
Factor 2	0.59443	-1.99454	0.01680	1.73970	-0.56156

Sumber: Hasil Olah Data (output SPSS)

Tabel 12. Iteration History

Iterasi	***Cluster***				
	1	**2**	**3**	**4**	**5**
1	0.843	0.223	0.184	0.443	0.636
2	0.000	0.203	0.072	0.000	0.000
3	0.000	0.179	0.069	0.000	0.000
4	0.000	0.119	0.072	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000

Sumber: Hasil Olah Data (output SPSS)

Berdasarkan tabel Iteration History, pada iterasi pertama masih terjadi perubahan pusat *cluster* yang relatif besar, terutama pada *Cluster* 1 (0,843) dan *Cluster* 5 (0,636). Nilai perubahan tersebut terus menurun pada iterasi berikutnya hingga seluruh *cluster* mencapai 0,000 pada iterasi kelima. Kondisi ini menunjukkan bahwa tidak terdapat lagi perpindahan pusat *cluster*, sehingga algoritma K-Means telah mencapai konvergensi. Dengan demikian, proses iterasi berhenti pada iterasi ke-5 dan lima *cluster* yang terbentuk merupakan hasil pengelompokan yang stabil.Jarak minimum antar pusat *cluster* yang terjadi dari hasil iterasi adalah 1.726.

Tabel 13. Final *Cluster* Centers

	Cluster				
	1	**2**	**3**	**4**	**5**
REGR Faktor Score 1 For Analysis 1	-2.07552	-0.21126	0.10236	0.15089	1.48721
REGR Faktor Score 2 For Analysis 1	-0.20436	-1.30614	0.17999	1.55584	-0.56812

Sumber: Hasil Olah Data (output SPSS)

Tabel diatas merupakan tabel *Final Cluster Centers*. Berdasarkan nilai pusat *cluster*, diperoleh pusat akhir (final *cluster* centers) menunjukkan posisi masing-masing *cluster* berdasarkan skor Faktor 1 dan Faktor 2 yang dihasilkan dari analisis faktor/PCA. Pada Faktor 1, *Cluster* 5 memiliki nilai pusat tertinggi (1,48721), diikuti *Cluster* 4 (0,15089), *Cluster* 3 (0,10236), *Cluster* 2 (-0,21126), dan *Cluster* 1 (-2,07552). Hal ini menunjukkan bahwa objek pada *Cluster* 5 memiliki karakteristik yang paling kuat terhadap variabel-variabel yang membentuk Faktor 1.

Dimana pada *Cluster* 1 memiliki karakteristik yang paling rendah. Sementara itu, pada Faktor 2, *Cluster* 4 memiliki nilai pusat tertinggi (1,55584), diikuti *Cluster* 3 (0,17999), sedangkan *Cluster* 5 (-0,56812), *Cluster* 1 (-0,20436), dan terutama *Cluster* 2 (-1,30614) memiliki nilai negatif. Perbedaan nilai pusat pada kedua faktor menunjukkan bahwa

setiap *cluster* memiliki karakteristik yang berbeda secara jelas, sehingga proses K-Means berhasil mengelompokkan objek ke dalam lima kelompok yang relatif homogen di dalam *cluster* dan heterogen antar*cluster* berdasarkan skor kedua faktor utama.

Tahapan selanjutnya yang perlu dilakukan yaitu melihat perbedaan variabel pada *cluster* yang terbentuk. Dalam hal ini dapat dilihat dari nilai F dan nilai probabilitas (sig.) masing-masing variabel, seperti pada tabel dibawah ini.

Tabel 14. Uji Anova K-Means

	Cluster		Error			
	Mean Square	**df**	**Mean Square**	**df**	**F**	**Sig.**
REGR factor score 1 for analysis 1	7.205	4	0.114	28	63.445	0.000
REGR factor score 2 for analysis 1	6.732	4	0.181	28	37.178	0.000

Sumber: Hasil Olah Data (output SPSS)

Berdasarkan hasil ANOVA diperoleh bahwa Faktor 1 dan Faktor 2 memiliki perbedaan rata-rata yang signifikan antar*cluster*. Hal ini ditunjukkan oleh nilai F sebesar 63,445 untuk Faktor 1 dan 37,178 untuk Faktor 2 dengan nilai signifikansi 0,000 ($p < 0,05$). Nilai F yang lebih besar pada Faktor 1 mengindikasikan bahwa Faktor 1 memberikan kontribusi yang lebih dominan dalam membedakan karakteristik antar*cluster* dibandingkan Faktor 2. Dengan demikian, kelima *cluster* yang terbentuk memiliki karakteristik yang berbeda secara statistik, sehingga hasil pengelompokan K-Means dapat dinilai baik dalam memisahkan objek berdasarkan kedua faktor utama.

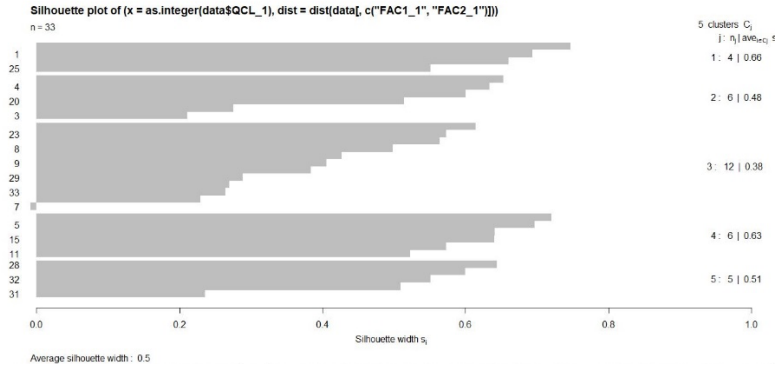
Tabel 15. Hasil Pengklasteran Indikator Sosial Ekonomi di Sumatera Utara Menggunakan K-Means

Cluster 1	Nias, Nias Selatan, Nias Utara, Nias Barat
Cluster 2	Mandailing Natal, Tapanuli Selatan, Tapanuli Tengah, Batu Bara, Padang Lawas Utara, Padang Lawas
Cluster 3	Labuhan Batu, Asahan, Simalungun, Langkat, Pakpak Bharat, Serdang Bedagai, Labuhanbatu Selatan, Labuanbatu Utara, Sibolga, Tanjungbalai, Tebing Tinggi, Gunungsitoli
Cluster 4	Tapanuli Utara, Toba, Dairi, Karo, Humbang Hasundutan, Samosir
Cluster 5	Deli Serdang, Pematangsiantar, Medan, Binjai, Padangsidimpuan

Validasi Hasil K-Means *Clustering*

Untuk mengetahui kualitas hasil pengelompokan yang terbentuk, dilakukan validasi internal menggunakan *Average Silhouette* Width dan *Davies–Bouldin Index* (DBI). Kedua ukuran ini digunakan untuk mengevaluasi tingkat kekompakan (compactness) setiap *cluster* serta tingkat pemisahan (separation) antar*cluster*.

Berdasarkan hasil perhitungan, diperoleh *Average Silhouette Width* sebesar 0,496. Nilai tersebut menunjukkan bahwa hasil pengelompokan memiliki struktur *cluster* yang cukup baik (*reasonable cluster structure*). Hal ini mengindikasikan bahwa sebagian besar kabupaten/kota memiliki tingkat kemiripan yang lebih tinggi terhadap anggota dalam *cluster* yang sama dibandingkan dengan anggota pada *cluster* lainnya. Meskipun masih terdapat beberapa objek yang berada di dekat batas antar*cluster*, secara umum hasil pengelompokan telah mampu membentuk kelompok yang cukup jelas sehingga layak digunakan untuk analisis lebih lanjut.



Gambar 3. *Silhouette Plot*

Berdasarkan Gambar 3, menunjukkan *Silhouette Plot* hasil pengelompokan menggunakan metode K-Means yang

terdiri atas 5 *cluster* dengan jumlah objek sebanyak 33 kabupaten/kota. Setiap batang merepresentasikan nilai *silhouette* suatu objek, sedangkan panjang batang menunjukkan tingkat kesesuaian objek terhadap *cluster* tempatnya berada. Semakin besar nilai *silhouette* (mendekati 1), semakin baik objek tersebut merepresentasikan *cluster*nya. Sebaliknya, nilai *silhouette*yang mendekati 0 menunjukkan bahwa objek berada di dekat batas antar*cluster*, sedangkan nilai negatif mengindikasikan kemungkinan objek lebih sesuai berada pada *cluster* lain.

Berdasarkan Gambar 3, seluruh objek memiliki nilai *silhouette* positif, sehingga tidak terdapat kabupaten/kota yang memiliki indikasi lebih sesuai ditempatkan pada *cluster* lain. Hal ini menunjukkan bahwa hasil pengelompokan telah menghasilkan pemisahan *cluster* yang cukup baik.

Selanjutnya, hasil perhitungan *Davies–Bouldin Index* (DBI) menunjukkan nilai sebesar 0,638. Nilai DBI yang lebih kecil menunjukkan kualitas *cluster* yang semakin baik karena memiliki tingkat kekompakan yang tinggi dan pemisahan antar*cluster* yang semakin jelas. Nilai DBI sebesar 0,638 yang berada di bawah satu mengindikasikan bahwa sebagian besar objek memiliki tingkat kemiripan yang lebih tinggi dengan anggota dalam *cluster* yang sama dibandingkan dengan anggota pada *cluster* lainnya. Dengan demikian, hasil pengelompokan menggunakan metode *K-Means* dapat dikatakan memiliki kualitas yang baik.

Berdasarkan kedua ukuran validasi tersebut, yaitu *Average Silhouette Width* sebesar 0,496 dan *Davies–Bouldin Index* sebesar 0,638, dapat disimpulkan bahwa hasil pengelompokan kabupaten/kota di Provinsi Sumatera Utara berdasarkan indikator pembangunan manusia memiliki kualitas yang memadai. Oleh karena itu, hasil *clustering* dinilai cukup representatif untuk digunakan dalam penyusunan profil karakteristik setiap *cluster* serta sebagai dasar interpretasi kondisi indikator pembangunan manusia di masing-masing kelompok wilayah.

Profil *Cluster*

Tabel 16. Profil Karakteristik Setiap *Cluster*

Cluster	Jumlah	HLS	RLS	AHH	TPT	Penduduk Miskin	Pengeluaran per Kapita
1	4	13,10	6,64	71,90	2,29	19,00	6.942
2	6	13,50	9,12	71,80	5,80	9,24	10.619
3	12	13,20	9,39	73,60	4,89	9,44	11.478
4	6	13,50	10,00	74,10	1,34	8,73	11.029
5	5	14,30	11,20	74,20	7,92	6,06	12.933

Interpretasi Profil *Cluster*

Berdasarkan hasil K-Means *Clustering*, diperoleh lima *cluster* yang memiliki karakteristik berbeda berdasarkan nilai rata-rata Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), Angka Harapan Hidup (AHH), Tingkat Pengangguran Terbuka (TPT), persentase penduduk miskin, dan pengeluaran per kapita.

***Cluster* 1**

Cluster 1 terdiri atas 4 kabupaten/kota dengan rata-rata HLS sebesar 13,10 tahun, RLS 6,64 tahun, AHH 71,90 tahun, TPT 2,29%, persentase penduduk miskin 19,00%, dan pengeluaran per kapita sebesar Rp6.942 ribu.

Cluster ini memiliki RLS terendah, persentase penduduk miskin tertinggi, dan pengeluaran per kapita terendah dibandingkan *cluster* lainnya. Hal tersebut menunjukkan bahwa daerah pada *cluster* ini masih menghadapi tantangan dalam aspek pendidikan dan kesejahteraan ekonomi. Meskipun tingkat penganggurannya relatif rendah, kondisi tersebut belum mampu diikuti oleh tingkat pendidikan dan kesejahteraan masyarakat yang baik. Oleh karena itu, *Cluster* 1 dapat dikategorikan sebagai wilayah dengan tingkat pembangunan manusia dan kondisi ekonomi relatif rendah.

***Cluster* 2**

Cluster 2 terdiri atas 6 kabupaten/kota dengan rata-rata HLS sebesar 13,50 tahun, RLS 9,12 tahun, AHH 71,80 tahun, TPT 5,80%, persentase penduduk miskin 9,24%, dan pengeluaran per kapita sebesar Rp10.619 ribu.

Dibandingkan *Cluster* 1, *cluster* ini menunjukkan kondisi pendidikan dan ekonomi yang lebih baik. Namun demikian, *cluster* ini memiliki TPT yang relatif tinggi, sehingga mengindikasikan bahwa peningkatan kualitas pendidikan belum sepenuhnya diikuti oleh penyerapan tenaga kerja yang optimal. Dengan demikian, *Cluster* 2 dapat dikategorikan sebagai wilayah dengan tingkat pembangunan manusia sedang, tetapi masih menghadapi permasalahan pada aspek ketenagakerjaan.

***Cluster* 3**

Cluster 3 merupakan *cluster* dengan jumlah anggota terbanyak, yaitu 12 kabupaten/kota. *Cluster* ini memiliki rata-rata HLS sebesar 13,20 tahun, RLS 9,39 tahun, AHH 73,60 tahun, TPT 4,89%, persentase penduduk miskin 9,44%, dan pengeluaran per kapita sebesar Rp11.478 ribu.

Secara umum, nilai rata-rata seluruh indikator berada pada tingkat menengah. Nilai AHH dan pengeluaran per kapita relatif lebih baik dibandingkan *Cluster* 2, sedangkan tingkat kemiskinan dan TPT masih berada pada tingkat sedang. Kondisi tersebut menunjukkan bahwa daerah-daerah pada *Cluster* 3 memiliki karakteristik pembangunan yang relatif seimbang tanpa adanya indikator yang sangat menonjol. Oleh karena itu, *Cluster* 3 dapat dikategorikan sebagai wilayah dengan tingkat pembangunan manusia menengah.

***Cluster* 4**

Cluster 4 terdiri atas 6 kabupaten/kota dengan rata-rata HLS sebesar 13,50 tahun, RLS 10,00 tahun, AHH 74,10 tahun, TPT 1,34%, persentase penduduk miskin 8,73%, dan pengeluaran per kapita sebesar Rp11.029 ribu. *Cluster* ini memiliki TPT terendah dibandingkan seluruh *cluster*, serta nilai RLS dan AHH yang relatif tinggi. Hal tersebut menunjukkan bahwa daerah-daerah pada *cluster* ini memiliki kualitas sumber daya manusia yang baik dan kondisi pasar kerja yang relatif stabil. Selain itu, tingkat kemiskinan juga relatif rendah sehingga *Cluster* 4 dapat dikategorikan sebagai wilayah dengan pembangunan manusia yang baik serta kondisi ketenagakerjaan yang kondusif.

***Cluster* 5**

Cluster 5 terdiri atas 5 kabupaten/kota dengan rata-rata HLS sebesar 14,30 tahun, RLS 11,20 tahun, AHH 74,20 tahun, TPT 7,92%, persentase penduduk miskin 6,06%, dan pengeluaran per kapita sebesar Rp12.933 ribu. *Cluster* ini memiliki HLS, RLS, AHH, dan pengeluaran per kapita tertinggi, serta persentase penduduk miskin terendah dibandingkan *cluster* lainnya. Kondisi tersebut menunjukkan bahwa daerah pada *cluster* ini memiliki kualitas pembangunan manusia dan tingkat kesejahteraan yang paling baik. Namun demikian, *cluster* ini juga memiliki TPT tertinggi, yang mengindikasikan bahwa daerah-daerah dalam *cluster* ini masih menghadapi tantangan pada aspek penyerapan tenaga kerja. Fenomena tersebut umum dijumpai pada wilayah perkotaan atau daerah dengan aktivitas ekonomi yang tinggi, di mana tingginya jumlah angkatan kerja dan mobilitas penduduk dapat menyebabkan tingkat pengangguran terbuka lebih tinggi dibandingkan daerah lainnya. Oleh karena itu, *Cluster* 5 dapat dikategorikan sebagai wilayah dengan pembangunan manusia sangat baik, tetapi menghadapi tantangan pada aspek ketenagakerjaan.

Secara keseluruhan, hasil K-Means menunjukkan adanya perbedaan karakteristik pembangunan manusia antar*cluster*. *Cluster* 1 didominasi oleh daerah dengan kondisi pendidikan dan kesejahteraan yang masih rendah. *Cluster* 2 dan *Cluster* 3 merepresentasikan wilayah dengan tingkat pembangunan menengah, meskipun *Cluster* 2 menghadapi tantangan yang lebih besar pada aspek ketenagakerjaan. *Cluster* 4 menunjukkan kondisi pembangunan manusia yang baik dengan tingkat pengangguran yang paling rendah, sedangkan *Cluster* 5 merupakan kelompok wilayah dengan tingkat pembangunan manusia tertinggi yang ditandai oleh indikator pendidikan, kesehatan, dan kesejahteraan ekonomi yang paling baik, namun masih memiliki tingkat pengangguran terbuka yang relatif tinggi

Kesimpulan

Berdasarkan analisis Principal Component Analysis (PCA) dan K-Means Clustering terhadap enam indikator pembangunan manusia pada 33 kabupaten/kota di Provinsi Sumatera Utara, kelayakan data teruji melalui nilai KMO sebesar 0,669 ($p < 0,001$) serta seluruh variabel yang memenuhi syarat MSA (> 0,50). Dua komponen utama yang terbentuk mampu menjelaskan 72,793% keragaman data. Komponen pertama didominasi oleh variabel rata-rata lama sekolah, pengeluaran per kapita, dan persentase penduduk miskin, sedangkan komponen kedua dibentuk oleh tingkat pengangguran terbuka dan angka harapan hidup.

Melalui metode Silhouette, diperoleh lima cluster optimal yang stabil pada iterasi kelima dan terbukti berbeda secara signifikan berdasarkan uji ANOVA ($p < 0,001$). Pengelompokan ini menghasilkan variasi karakteristik wilayah, mulai dari Cluster 1 (pembangunan dan kesejahteraan relatif rendah), Cluster 2 dan 3 (pembangunan menengah), Cluster 4 (kualitas pembangunan baik dengan pengangguran terendah), hingga Cluster 5 (kualitas pembangunan terbaik pada indikator pendidikan, kesehatan, dan pengeluaran, meski pengangguran masih tinggi). Evaluasi validasi internal menunjukkan Average Silhouette Width sebesar 0,496 dan Davies–Bouldin Index sebesar 0,638, menegaskan bahwa kombinasi PCA dan K-Means efektif mereduksi dimensi sekaligus mengelompokkan wilayah secara akurat sebagai dasar kebijakan pembangunan terfokus.

Ucapan Terima Kasih

Terimakasih kepada ibu Dr. Rina Filia Sari, M.Si selaku dosen pengampu matakuliah analisis multivariat dan terimakasih kepada seluruh anggota kelompok atas partisipasinya dalam penulisan artikel ini.

Daftar Pustaka

[1] D. Hidayat, N. Ratnawati, and Syafri, "Pengaruh Indeks Pembangunan Manusia dan Jumlah Penduduk Terhadap Produk Regional Domestik Bruto di Provinsi Riau Tahun 2010-2022," *J. Ekon. KIAT*, vol. 36, no. 1, pp. 24–31, 2025, doi: 10.25299/kiat.2025.23268.

[2] S. B. Loklomin, "Pemodelan Indeks Pembangunan Manusia Di Kepulauan Maluku Dengan Pendekatan Estimasi Interval Parameter Model Regresi Semiparametrik Spline Truncated," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 13, no. 2, pp. 125–134, 2019, doi: 10.30598/barekengvol13iss2pp125-134ar820.

[3] Y. Yulianti and S. Qomariah, "Indeks Pembangunan Manusia Ilmu Pengetahuan," *CENDEKIA J. Ilmu Sos. Bhs. dan Pendidik.*, vol. 5, no. 1, pp. 190–200, 2025, doi: https://doi.org/10.55606/cendikia.v5i1.3512.

[4] Hanniva, A. Kurnia, S. Rahardiantoro, and A. A. Mattjik, "Penggerombolan Kabupaten/Kota di Indonesia Berdasarkan Indikator Indeks Pembangunan Manusia Menggunakan Metode K-Means dan fuzzy C-Means" *Xplore J. Stat.*, vol. 11, no. 01, pp. 36–47, 2022, doi: 10.29244/xplore.v11i1.855.

[5] N. Nurhasanah, N. Salwa, L. Ornila, A. Hasan, and M. Mardani, "Classifying regencies and cities on human development index dimensions : Application of K-Means cluster analysis," *Sains Sosio Hum.*, vol. 5, no. 2, pp.

759–765, 2021, doi: https://doi.org/10.22437/jssh.v5i2.15753.
[6] M. Valentino, Y. Sipahutar, and M. Farhan, "Pembangunan Global Menggunakan Metode Pca Dan Clustering K-Means Tahun 2000-2020," *Ilmu Komput. dan Sist. Inf.*, vol. 13, no. 2, pp. 1–16, 2025, doi: https://doi.org/10.24912/jiksi.v13i2.35130.
[7] R. S. Tobing, O. H. Sigalingging, R. K. Sinaga, and A. Lubis, "Penerapan Principal Component Analysis (PCA) untuk Reduksi Dimensi dan Pemetaan Karakteristik Nutrisi pada Produk Makanan Kemasan di Indonesia dengan memilih makanan kemasan yang lebih sehat dan lebih baik sehingga dapat mencegah Makanan Kemasan dan Tan," *Mat. Ilmu Pengetah. Alam, Kebumian dan Angkasa*, vol. 4, no. 1, pp. 20–30, 2026, doi: https://doi.org/10.62383/algoritma.v4i1.891.
[8] F. H. Lubis, S. F. Ashar, O. M. F. A. M. S, and A. B. Amari, "Studi Pengelompokan Multimetode Provinsi di Sumatera Utara Menggunakan Pendekatan PCA dan K-Means," *Algoritm. J. Ilmu Komput. Inform.*, vol. 9, no. 2, pp. 72–83, 2025, doi: 10.30829/algoritma.v9i2.25017.
[9] R. Handoyo, R. R. M, and S. M. Nasution, "Perbandingan Metode Clustering Menggunakan Metode Single Linkage Dan K - Means Pada Pengelompokan Dokumen," vol. 15, no. 2, pp. 73–82, 2014.
[10] R. Kurniawan, M. S. Hasibuan, and R. Hasibuan, "Klasterisasi Wilayah Prioritas Vaksin Menggunakan Algoritma K-Means Clustering," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 3, pp. 1585–1592, 2023, doi: 10.30865/klik.v4i3.1334.
[11] H. Sibarani, Solikhun, W. Saputra, I. Gunawan, and Z. M. Nasution, "Penerapan Metode K-Means Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Sumatera Utara Berdasarkan Indikator Indeks Pembangunan Manusia," *J. Mhs. Tek. Inform.*, vol. 6, no. 1, pp. 154–161, 2022, doi: https://doi.org/10.36040/jati.v6i1.4590.
[12] N. Wahyunisari and R. Kurniawan, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Data Penjualan Pakaian (Studi Kasus : Umkm Lima Media Kuningan)," *J. Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1398–1403, 2024.
[13] N. F. Setyawati *et al.*, *Metodologi Riset Kesehatan*, 1st ed. Purbalingga: CV. Eureka Media Aksara, 2023.
[14] D. Hediyati and I. M. Suartana, "Penerapan Principal Component Analysis (PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian Di Kabupaten Bojonegoro," *J. Inf. Eng. Educ. Technol.*, vol. 05, no. 2, pp. 49–54, 2021.
[15] E. Novia and P. Hendikawati, "Average Linkage Hierarchical Cluster untuk Pengelompokan Kabupaten / Kota di Provinsi Jawa Tengah Berdasarkan Variabel Pencegahan Terjadinya Stunting," in *PRISMA, Prosiding Seminar Nasional Matematika*, Jurusan Matematika, Universitas Negeri Semarang 1., 2024, pp. 702–711. [Online]. Available: https://proceeding.unnes.ac.id/prisma
[16] M. Yafi, R. Goejantoro, and A. T. R. Dani, "K-Medoids Algorithm Clustering With Principal Component Analysis (PCA) (Case Study: Districts/Cities On The Borneo Island Based On Poverty Indicators)," *Statistika*, vol. 11, no. 2, pp. 31–43, 2023, doi: 10.14710/JSUNIMUS.11.2.2023.31-43.
[17] L. Izzah and A. Jananto, "Penerapan Algoritma K-Means Clustering Untuk Perencanaan Kebutuhan Obat Di Klinik Citra Medika," *Progresif J. Ilm. Komput.*, vol. 18, no. 1, pp. 69–76, 2022.
[18] Purnomo *et al.*, *Analisis Data Multivariat*, vol. 2. 2022.
[19] Ameliya *et al.*, "Penerapan Principal Component Analysis untuk Menentukan Faktor- Faktor yang Mempengaruhi Kemiskinan di Sumatera Utara yang Mempengaruhi Kemiskinan di Papua dengan Principal Component Analysis "," *Mat. Ilmu Pengetah. Alam, Kebumian dan Angkasa*, vol. 4, no. 1, pp. 01–19, 2026.
[20] R. F. Purba, A. A. Panjaitan, L. T. Butar-butar, J. F. E. D. L. Sitorus, R. Rumapea, and I. M. S. S, "Implemntasi K-Means Clustering Untuk Menentukan Tingkat Bencana Rawan Banjir Di Wilayah Sumatera Utara," *J. Tugas Akhir Manaj. Inform. Komputerisasi Akunt.*, vol. 4, no. 2, pp. 210–215, 2024, doi: https://doi.org/10.46880/tamika.Vol4No2(SEMNASTIK).pp210-215.