Jurnal Sistem Komputer dan Informatika (JSON)
Volume 6, Nomor 4, Juni 2025
e-ISSN 2685-998X
DOI 10.30865/json.v6i4

Segmentasi Provinsi di Indonesia Berdasarkan Akses Fasilitas Dasar dan Pengeluaran Rumah Tangga Menggunakan K-Means

Devi Saputri, Hindayati Mustafidah*, Feri Wibowo, Dimara Kusuma Hakim
Fakultas Teknik dan Sains, Prodi Teknik Informatika, Universitas Muhammadiyah Purwokerto, Banyumas 53182, Indonesia
Email: [1]saputrid488@gmail.com, [2]h.mustafidah@ump.ac.id, [3]feriwibowo@ump.ac.id, [4]dimarakusumahakim@gmail.com
Email Penulis Korespondensi: h.mustafidah@ump.ac.id*
Submitted: **20/05/2025**; Accepted: **10/06/2025**; Published: **30/06/2025**

Abstrak– Pemerataan akses terhadap fasilitas dasar dan peningkatan kesejahteraan masyarakat di Indonesia masih menjadi tantangan besar, khususnya antarprovinsi. Meskipun dalam satu dekade terakhir telah terjadi kemajuan pembangunan, ketimpangan antarwilayah masih tampak nyata. Ketimpangan ini tercermin dari perbedaan signifikan dalam akses terhadap air minum layak, sanitasi, listrik, tempat tinggal yang layak, serta pengeluaran rumah tangga per kapita per bulan. Penelitian ini bertujuan untuk mengelompokkan 34 provinsi di Indonesia berdasarkan indikator akses terhadap fasilitas dasar dan pengeluaran rumah tangga guna mengidentifikasi pola ketimpangan pembangunan wilayah. Metode yang digunakan adalah algoritma K-Means Clustering dengan variabel mencakup kepemilikan rumah, akses air minum layak, sanitasi, listrik, penggunaan gas, serta pengeluaran rumah tangga yang berkaitan dengan fasilitas tersebut. Hasil segmentasi menunjukkan terbentuknya dua klaster: Klaster 1 terdiri dari 29 provinsi dengan akses yang lebih baik terhadap fasilitas dasar dan tingkat pengeluaran rumah tangga yang lebih tinggi, namun dengan tingkat kepemilikan rumah yang relatif lebih rendah. Klaster 2 mencakup 5 provinsi dengan akses terbatas terhadap infrastruktur dasar dan tingkat pengeluaran yang lebih rendah, namun dengan tingkat kepemilikan rumah yang lebih tinggi. Temuan ini memberikan gambaran mengenai ketimpangan pembangunan antarprovinsi di Indonesia yang dapat menjadi acuan bagi pemerintah dalam perumusan kebijakan pembangunan wilayah yang lebih merata.

Kata Kunci: Fasilitas dasar; Ketimpangan Wilayah; K-Means; pengeluaran rumah tangga; Segmentasi.

Abstract– Ensuring equitable access to basic facilities and improving community welfare remain major challenges in Indonesia, particularly across provinces. Although there has been development progress over the past decade, regional disparities are still evident. These inequalities are reflected in significant differences in access to safe drinking water, sanitation, electricity, adequate housing, and monthly household expenditure per capita. This study aims to cluster 34 provinces in Indonesia based on indicators of access to basic facilities and household expenditures to identify patterns of regional development disparities. The method used is the K-Means Clustering algorithm, with variables including home ownership, access to safe drinking water, sanitation, electricity, gas usage, and household expenditures related to these facilities. The segmentation results show the formation of two clusters: Cluster 1 consists of 29 provinces with better access to basic facilities and higher household expenditure levels but relatively lower home ownership. Cluster 2 includes 5 provinces with limited access to basic infrastructure and lower expenditure levels but higher home ownership rates. These findings provide an overview of inter-provincial development disparities in Indonesia and can serve as a reference for the government in formulating more regionally targeted development policies.

Keywords: Basic Facilities; Household Expenditures; K-Means; Regional Disparity; Segmentation.

1. PENDAHULUAN

Pemerataan akses terhadap fasilitas dasar dan peningkatan kesejahteraan masyarakat masih menjadi tantangan utama yang dihadapi oleh banyak negara, termasuk Indonesia [1]. Pemerataan pembangunan dan peningkatan kesejahteraan masyarakat merupakan isu strategis yang terus menjadi perhatian dalam agenda pembangunan nasional Indonesia. Meskipun telah terjadi berbagai kemajuan pembangunan dalam beberapa dekade terakhir, disparitas antarwilayah, khususnya antara kawasan barat dan timur Indonesia, masih sangat mencolok [2]. Ketimpangan ini tercermin dari perbedaan signifikan dalam hal akses terhadap fasilitas dasar seperti air minum layak, sanitasi yang memadai, listrik dari PLN, bahan bakar gas, serta ketersediaan tempat tinggal yang layak[3]**.** Selain itu, terdapat pula ketimpangan dalam kemampuan ekonomi masyarakat yang dapat diukur melalui indikator pengeluaran rumah tangga. Ketidakseimbangan yang terjadi mengakibatkan pembentukan wilayah-wilayah yang terbelakang dan miskin, sehingga menjadi kendala dalam pencapaian tujuan pembangunan berkelanjutan (*Sustainable Development Goals*/SDGs), khususnya pada target ke-6 (penyediaan air bersih dan sanitasi), taget ke-7 (energi yang bersih dan terjangkau), serta target ke-11 (pembangunan kota dan permukiman yang inklusif, aman, tangguh, dan berkelanjutan) serta pemerataan kesejahteraan di seluruh wilayah Indonesia [4]–[6].

Dalam kerangka pembangunan berkelanjutan, pembangunan infrastruktur dasar berperan penting dalam mendorong pertumbuhan ekonomi serta meningkatkan kualitas hidup masyarakat [7]. Ketersediaan fasilitas dasar seperti air bersih, sanitasi yang memadai, listrik, dan bahan bakar gas merupakan kebutuhan esensial yang memengaruhi kualitas hidup masyarakat. Ketersediaan dan kemudahan akses terhadap fasilitas tersebut berperan penting dalam menentukan kualitas hidup serta derajat kesehatan masyarakat [8]. Sementara itu, indikator ekonomi rumah tangga seperti pengeluaran bulanan dapat digunakan untuk mengidentifikasi kemampuan daya beli masyarakat, kondisi sosial-ekonomi, serta distribusi kesejahteraan di suatu wilayah [9]. Oleh karena itu, Kombinasi antara kedua aspek ini yaitu akses fasilitas dasar dan kondisi ekonomi rumah tangga dapat menjadi fondasi dalam melakukan pemetaan wilayah berdasarkan tingkat kesejahteraan secara lebih menyeluruh

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 6, Nomor 4, Juni 2025
e-ISSN 2685-998X
DOI 10.30865/json.v6i4

menggunakan pendekatan berbasis data untuk mengidentifikasi wilayah-wilayah yang memerlukan intervensi prioritas. Salah satu pendekatan tersebut adalah segmentasi wilayah berbasis karakteristik pengeluaran rumah tangga dan infrastruktur dasar, yang bertujuan untuk mengelompokkan wilayah ke dalam klaster dengan karakteristik serupa, guna mendukung perumusan kebijakan yang lebih tepat sasaran.

Seiring dengan perkembangan teknologi dan meningkatnya kebutuhan akan analisis data yang efisien, pendekatan berbasis *Artificial Intelligence* (AI) semakin banyak digunakan dalam eksplorasi data kompleks dan pengambilan keputusan. Salah satu cabang dari AI adalah *Machine Learning* (ML), yang memungkinkan sistem mempelajari pola dalam data tanpa arahan eksplisit [10]. Dalam ML, terdapat pendekatan *unsupervised learning* yang memungkinkan analisis struktur data tanpa label, salah satunya melalui metode *clustering* [11], [12]. K-Means adalah salah satu algoritma *clustering* populer dalam *data mining* yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik, sehingga memudahkan dalam menemukan pola-pola tersembunyi [13]–[15]. Algoritma ini bekerja dengan cara meminimalkan jarak antar data dalam satu klaster sekaligus memperbesar jarak antar klaster yang berbeda untuk menghasilkan pengelompokan yang optimal, dengan kelebihannya yaitu kesederhanaan, efisiensi komputasi yang cepat, serta kemampuannya dalam menangani dataset berskala besar. [16]–[18]. K-Means sangat cocok digunakan dalam proses segmentasi wilayah berdasarkan data numerik, seperti persentase akses fasilitas dasar dan pengeluaran rumah tangga perkapita perbulan.

Berbagai penelitian sebelumnya telah menerapkan metode machine learning dalam analisis *clustering*. Beberapa penelitian menggunakan K-Nearest Neighbor (KNN) untuk mengelompokkan objek, seperti penelitian yang berhasil mengelompokkan ikan air tawar dengan akurasi 70% [19]. Hierarchical Clustering diterapkan dalam 2 penelitian berbeda, yaitu pada analisis infrastruktur jalan serta pada tingkat kemiskinan, yang keduanya menghasilkan tiga klaster utama [20], [21]. K-Medoids digunakan untuk pengelompokan wilayah di Papua berdasarkan indikator kemiskinan menghasilkan empat klaster, sedangkan Fuzzy C-Means diterapkan di Jawa Tengah untuk mengelompokkan tingkat kemiskinan, menghasilkan lima klaster [22] [23]. Random forest juga digunakan dalam prediksi pengeluaran rumah tangga berbasis data survei dan geospasial di Kamboja, dengan hasil presisi tertinggi [24]. Selanjutnya, penelitian [25] menekankan pentingnya akses terhadap sanitasi, listrik, dan infrastruktur jalan dalam menurunkan tingkat kemiskinan.

Metode K-Means juga banyak digunakan dalam segmentasi wilayah untuk mengidentifikasi pola-pola kemiskinan dan kesejahteraan di berbagai daerah. Misalnya, pengelompokkan kabupaten/kota di Sulawesi Selatan dilakukan berdasarkan indikator IPM untuk meningkatkan angka pembangunan manusia menjadi tiga klaster [14], mengelompokkan kabupaten/kota berdasarkan indikator kemiskinan di Bangka Belitung menjadi tiga klaster [17], dan enam klaster di Sumatera Utara [26]. Pengelompokkan rumah tangga berdasarkan status kepemilikan rumah menghasilkan 3 klaster [27], Sementara itu, indikator rumah layak huni juga telah dikelompokkan dalam beberapa penelitian dan masing-masing menghasilkan 4 klaster [12], [28]. Lebih lanjut, penelitian lainnya menunjukkan bahwa mengelompokkan provinsi di Indonesia menggunakan K-Means menggunakan kombinasi antara faktor ekonomi dan lingkungan pada tahun 2022 berhasil mengidentifikasi 3 klaster yang mencerminkan keberagaman kondisi sosial ekonomi serta memberikan wawasan strategis dalam perumusan kebijakan pembangunan daerah [9].

Meski demikian, sebagian besar penelitian sebelumnya masih berfokus pada satu dimensi saja, baik akses terhadap fasilitas dasar maupun aspek ekonomi rumah tangga, tanpa mengintegrasikan keduanya secara komprehensif. Padahal, kombinasi kedua variabel tersebut penting untuk memberikan gambaran yang lebih menyeluruh mengenai kondisi kesejahteraan dan pemerataan pembangunan di suatu wilayah. Penelitian ini menitikberatkan pada analisis pemerataan akses terhadap fasilitas dasar yang merupakan kebutuhan esensial dan sangat memengaruhi kualitas hidup serta kesehatan masyarakat. Sementara itu, pengeluaran rumah tangga per kapita per bulan digunakan sebagai representasi kemampuan ekonomi masyarakat dalam memenuhi kebutuhan hidupnya. Oleh karena itu, penelitian ini bertujuan melakukan segmentasi terhadap 34 provinsi di Indonesia menggunakan algoritma K-Means, dengan variabel akses fasilitas dasar (seperti rumah layak huni, air bersih, listrik, gas, dan sanitasi) serta pengeluaran rumah tangga per kapita per bulan sebagai representasi kesejahteraan ekonomi. Validitas jumlah klaster akan dianalisis menggunakan *Elbow method* dan *Silhouette score* untuk memastikan kualitas pengelompokan. Hasil segmentasi diharapkan dapat memberikan gambaran yang lebih jelas mengenai kesenjangan pembangunan di Indonesia dan menjadi acuan dalam perumusan kebijakan pembangunan yang lebih merata dan berkelanjutan sesuai kebutuhan wilayah.

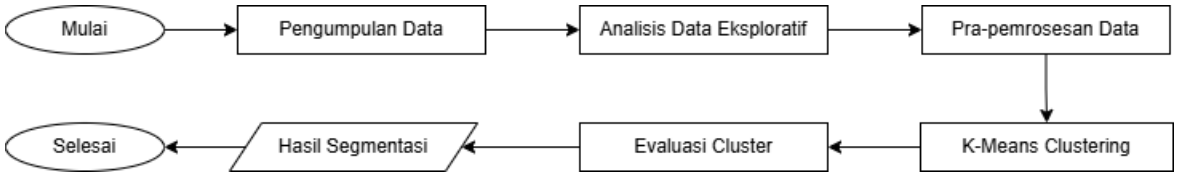
2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan dalam penelitian ini meliputi beberapa proses, dimulai dari pengumpulan data, analisis data eksploratif, pra-pemrosesan data, hingga penerapan algoritma K-Means untuk segmentasi wilayah. Seluruh proses analisis dilakukan menggunakan bahasa pemrograman Python karena Python menyediakan berbagai pustaka *(library)* yang mendukung pengolahan data dan implementasi machine learning, seperti *numpy* untuk operasi numerik,

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 6, Nomor 4, Juni 2025
e-ISSN 2685-998X
DOI 10.30865/json.v6i4

pandas untuk manipulasi data, *scikit-learn* untuk penerapan algoritma K-Means dan evaluasi klaster, serta *matplotlib* dan *seaborn* untuk visualisasi data [15]. Alur tahapan penelitian ditampilkan pada Gambar 1.



Gambar 1. Alur Tahapan Penelitian

2.2 Pengumpulan Data

Sumber data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari publikasi resmi Badan Pusat Statistik (BPS) tahun 2023. Informasi yang terkandung dalam publikasi ini mencakup perumahan dan kesehatan lingkungan provinsi di Indonesia [6]. Dataset yang digunakan mencakup 34 provinsi dan terdiri atas 12 atribut yang dikelompokkan ke dalam dua kategori utama, yaitu fasilitas dasar dan pengeluaran rumah tangga. Data tersebut diinput secara manual ke dalam Microsoft Excel dan kemudian dikonversi ke dalam format CSV.

2.3 Analisis data eksploratif

Analisis eksploratif dilakukan untuk memperoleh pemahaman awal mengenai karakteristik dataset yang digunakan [26]. Tahapan ini mencakup identifikasi nilai-nilai yang hilang (*null values*) serta pendeteksian data duplikat untuk memastikan kelengkapan data dan mengurangi kemungkinan redundansi yang dapat memengaruhi hasil analisis lebih lanjut

2.4 Pra-pemrosesan data

Pra-pemrosesan data dilakukan sebagai tahap awal sebelum penerapan algoritma K-Means, mencakup penanganan nilai hilang, penghapusan data duplikat, serta normalisasi menggunakan *Robust Scaler*. Seluruh variabel dikonversi ke dalam format yang seragam untuk memastikan kesetaraan skala antar fitur pada proses analisis lanjutan [29]. *Robust Scaler* dipilih karena kemampuannya dalam meminimalkan pengaruh outlier [30]. Proses normalisasi dilakukan menggunakan persamaan (1).

$$X_{scaled} = \frac{X - Median(X)}{IQR(X)} \quad (1)$$

Dimana:
X_{scaled} = Nilai yang telah di normalisasi,
X = Nilai data asli,
$Median(X)$ = Median dari data asli,
$IQR(X)$ = Rentang antar kuartil Q3 – Q1, dimana Q1 adalah kuartil pertama (25%) dan Q3 adalah kuartil ketiga (75%).

2.5 K-Means

Algoritma K-Means diterapkan untuk membagi provinsi ke dalam kelompok-kelompok yang memiliki kemiripan karakteristik berdasarkan atribut yang telah dipilih. Untuk menentukan jumlah klaster yang paling sesuai, digunakan metode *Elbow* yang menganalisis variasi total data pada berbagai nilai K dan menemukan titik di mana penurunan variasi mulai melambat. Perhitungan jarak antara setiap data dengan pusat klaster (*centroid*) dilakukan menggunakan rumus *Euclidean*, seperti yang ditunjukkan pada persamaan (3).
Proses algoritma K-Means terdiri atas beberapa langkah berikut:

1. Menentukan Jumlah klaster (K) menggunakan *Elbow method* yang bertujuan untuk menemukan jumlah klaster optimal dengan menganalisis penurunan *within-cluster sum of squares* (WCSS) terhadap berbagai nilai *K* [9]. Menentukan jumlah klaster dengan persamaan (2)

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

Dimana:
K = Jumlah klaster,
C_k = Klaster ke-k,
$x_i \in C_k$ = Data ke-1 yang termasuk dalam klaster ke-k,
μ_k = Titik pusat (*centroid*) klaster ke – k,
2. Menginisialisasi titik pusat klaster (*centroid*) secara acak untuk masing-masing klaster.
3. Pengelompokan data berdasarkan ke *centroid* terdekat. Jarak antara titik data dan *centroid* dihitung menggunakan rumus *Euclidean* pada persamaan (3) [31].

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 6, Nomor 4, Juni 2025
e-ISSN 2685-998X
DOI 10.30865/json.v6i4

$$d = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (3)$$

Dimana:
d = Jarak *Euclidean* antara dua titik data,
$x_i - y_i$ = Nilai Koordinat pada dimensi ke-i,
n = Banyaknya variabel / dimensi

4. Pembaruan *centroid* menggunakan rata-rata posisi data dalam klaster menggunakan persamaan (4):

$$\mu_k = \frac{1}{|C_k|}\sum_{x \epsilon C_k} x \quad (4)$$

Dimana:
μ_k = Titik pusat (*centroid*) klaster ke – k,
C_k = Kelompok data dalam klaster ke – k,
$|C_k|$ = Jumlah data klaster ke – k,
x = Titik data yang termasuk dalam klaster C_k.
5. Iterasi langkah 3 dan 4 hingga konvergensi atau jumlah iterasi maksimum tercapai.
6. Setelah konvergensi, hasil akhir adalah pengelompokan titik data ke dalam klaster yang telah ditentukan, dengan *centroid* yang merepresentasikan posisi rata-rata dari setiap kluster.

2.6 Evaluasi klaster

Evaluasi hasil klaster menggunakan *Indeks Silhouette* (SI), yang berfungsi untuk mengukur seberapa baik suatu data terletak dalam klasternya dibandingkan dengan klaster lainnya [32]. SI dihitung berdasarkan rata-rata jarak dalam-klaster (*a(i)*) dan jarak ke klaster terdekat (*b(i)*), sebagaimana ditunjukkan pada persamaan (5). Nilai SI berada pada rentang -1 hingga 1. Nilai yang mendekati 1 menunjukkan bahwa hasil pengelompokan data memiliki kualitas dan tingkat pemisahan yang baik antar klaster [33].

$$S(i) = \frac{C(j) - D(j)}{\max(C(j), D(j))} \quad (5)$$

Dimana:
C(j) = Rata-rata jarak dari titik ke-j ke semua titik lain dalam klaster yang sama,
D(j) = Rata-rata jarak dari titik ke-j ke titik klaster yang terdekat.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Tahap awal dalam proses analisis ini adalah pengumpulan data, di mana data yang telah dikonversi ke dalam format CSV diolah menggunakan bahasa pemrograman Python melalui platform Google Colab. Dataset mencakup 34 provinsi di Indonesia dan memuat variabel-variabel yang diklasifikasikan menjadi dua kategori utama. Kategori pertama adalah fasilitas dasar, yang terdiri atas enam variabel: kepemilikan rumah (X_1), akses air minum layak (X_2), akses sanitasi yang layak (X_3), akses listrik PLN (X_4), penggunaan gas sebagai bahan bakar utama(X_5), dan akses rumah layak huni(X_6). Kategori kedua mencakup pengeluaran rumah tangga, yang juga terdiri atas enam variabel: pengeluaran untuk listrik (X_7), air (X_8), bahan bakar (X_9), gas (X_{10}), pengeluaran kebutuhan rumah tangga lainnya(X_{11}) serta untuk sewa/kontrak(X_{12})). Setelah dikonversi dan dikategorikan, data tersebut kemudian digunakan dalam proses analisis klaster dengan algoritma K-Means. Ringkasan data dapat dilihat pada Tabel 1.

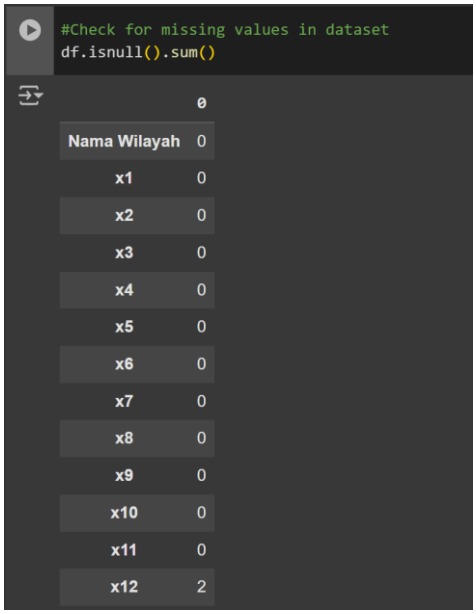
Tabel 1. Dataset fasilitas dasar dan pengeluaran rumah tangga

Nama Provinsi	X_1	X_2	X_3	X_4	….	X_{12}
Aceh	84,12	89,74	78,85	99,67	….	3059
Sumatera Utara	71,46	92,19	84,18	98,9	….	3516
Sumatera Barat	72,61	85,59	70,97	98,99	….	3542
Riau	77,56	90,47	84,58	95,54	….	4423
….	….	….	….	….	….	….
Maluku Utara	90,26	89,01	80,64	91,58	….	3553
Papua Barat	82,94	81,57	76,3	83,2	….	4059
Papua	85,31	66,49	43	47,07	….	4452

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 6, Nomor 4, Juni 2025
e-ISSN 2685-998X
DOI 10.30865/json.v6i4

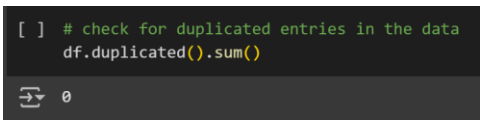
3.2 Analisis Data Eksploratif

Analisis data eksploratif dimulai dengan memeriksa keberadaan *null values*, kemudian dilanjutkan dengan pemeriksaan duplikasi data. Hasil eksplorasi awal menunjukkan adanya *null values* pada variabel X_{12} (Pengeluaran sewa/kontrak) di dua provinsi, yakni Sulawesi Tenggara dan Gorontalo seperti yang terlihat pada Gambar 2. *Null values* tersebut disebabkan tidak tersedianya data dari sumber asli untuk variabel tersebut di kedua wilayah tersebut. Oleh karena itu, kolom tersebut dihapus guna menghindari potensi bias dalam analisis lanjutan [34].



Gambar 2. Hasil Pengecekan Null values

Langkah berikutnya adalah pemeriksaan terhadap redundansi data, yaitu identifikasi keberadaan data ganda atau duplikat yang dapat memengaruhi hasil analisis. Berdasarkan hasil pengecekan sebagaimana ditunjukkan pada Gambar 3, tidak ditemukan adanya duplikasi data dalam dataset. Dengan demikian, dataset dapat dipastikan dalam kondisi bersih dan siap untuk tahap pra-pemrosesan selanjutnya.



Gambar 3. Hasil pengecekan duplikat data

3.3 Pra-pemrosesan Data

Tahap pra-pemrosesan data diawali dengan penghapusan data yang mengandung *null values* yang telah diidentifikasi sebelumnya pada Gambar 2. Selanjutnya, seluruh variabel numerik dalam dataset dinormalisasi menggunakan metode *Robust Scaler*. Pemilihan metode ini bertujuan untuk meminimalkan pengaruh *outlier* yang dapat mengganggu hasil pengelompokan, karena *Robust Scaler* menggunakan median dan *interquartile range (IQR)* sebagai dasar transformasi [30]. Hasil dari proses normalisasi ditampilkan dalam Tabel 2, langkah ini memastikan bahwa setiap variabel memiliki skala yang sebanding, sehingga tidak ada variabel yang mendominasi proses *clustering*.

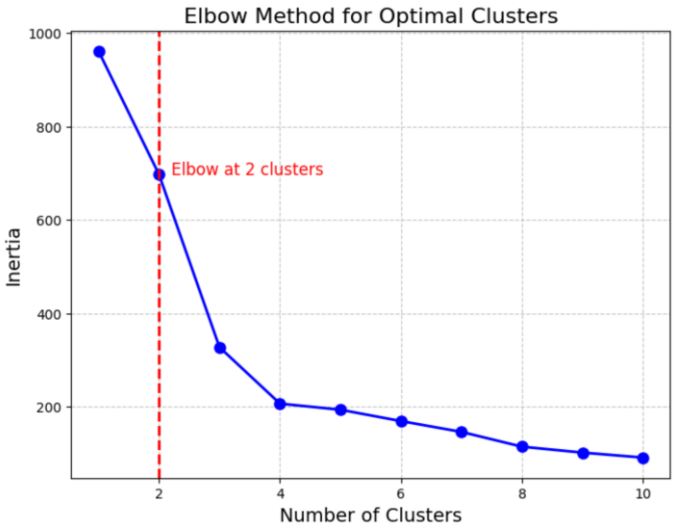
Tabel 2. Hasil Normalisasi Data menggunakan Robust Scaler

Nama Provinsi	X_1	X_2	X_3	X_4	….	X_{11}
Aceh	-0,1503	-0,0194	-0,6314	0,2623	….	-0,52938
Sumatera Utara	-1,7982	0,19214	0,1115	0,0795	….	-0,11298
Sumatera Barat	-1,6486	-0,3778	-1,7296	0,1009	….	-0,08929
Riau	-1,0042	0,04361	0,1672	-0,718	….	0,71344
….	….	….	….	….	….	….
Maluku Utara	0,64888	-0,0825	-0,3819	-1,658	….	-0,07927
Papua Barat	-0,3039	-0,725	-0,9868	-3,647	….	0,38178
Papua	0,00456	-2,0272	-5,6279	-12,22	….	0,73986

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 6, Nomor 4, Juni 2025
e-ISSN 2685-998X
DOI 10.30865/json.v6i4

3.4 Implementasi Algoritma K-Means

Langkah awal dalam penerapan algoritma K-Means adalah menentukan jumlah klaster yang optimal melalui metode *Elbow*. Metode ini berfungsi untuk menemukan titik di mana penurunan inertia mulai melambat secara signifikan, yang mengindikasikan jumlah klaster yang optimal. Evaluasi dilakukan dengan membandingkan nilai *inertia* pada rentang jumlah klaster dari 1 hingga 10. Hasil visualisasi metode *Elbow* ditampilkan pada Gambar 4.



Gambar 4. Grafik Metode *Elbow*

Hasil dari penerapan metode *Elbow* pada Gambar 4, menunjukkan bahwa titik optimal terletak pada jumlah 2 klaster, ditandai dengan penurunan signifikan yang melambat setelah klaster ke-2. Untuk memastikan kualitas hasil klaster, dilakukan evaluasi menggunakan *indeks Silhouette*. Pada Tabel 3, disajikan hasil evaluasi klaster yang diperoleh dengan menggunakan *indeks Silhouette* dengan rentang jumlah klaster 2 hingga 9, nilai tertinggi diperoleh saat jumlah klaster adalah dua, yaitu 0,593459, yang mengindikasikan kualitas pemisahan klaster yang cukup baik. Oleh karena itu, jumlah klaster yang digunakan dalam penelitian ini adalah dua klaster, berdasarkan hasil metode *Elbow* dan nilai *indeks Silhouette* tertinggi.

Tabel 3. Nilai *silhouette* pada setiap klaster

Jumlah Klaster	Nilai *Silhouette*
2	0,593459
3	0,487117
4	0,342203
5	0,212412
6	0,214376
7	0,203880
8	0,179779
9	0,158099

Setelah jumlah klaster ditentukan, langkah selanjutnya adalah menentukan titik pusat awal (*centroid*) yang ditampilkan pada Tabel 4. Penentuan *centroid* ini dilakukan menggunakan data yang telah dinormalisasi pada tahap sebelumnya (Tabel 2).

Tabel 4. Titik pusat klaster (*centroid*) pertama

Centroid	X_1	X_2	X_3	X_4	….	X_{11}
Centroid 1	0,7348	0,522	0,0474	0,2552	….	-0,0091
Centroid 2	0,7114	-0,1395	-1,0746	-2,5531	….	-1,2237

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 6, Nomor 4, Juni 2025
e-ISSN 2685-998X
DOI 10.30865/json.v6i4

Tabel 4 menunjukkan nilai awal *centroid* untuk masing-masing klaster. Setelah penentuan *centroid* awal, proses selanjutnya adalah melakukan pengelompokan data berdasarkan jarak *Euclidean* terhadap setiap *centroid*. Dalam penelitian ini, iterasi K-Means menunjukkan hasil konvergen pada iterasi kedua.

3.5 Hasil Segmentasi

Proses pengelompokkan dilakukan menggunakan algoritma K-Means yang diimplementasikan menggunakan bahasa pemrograman Python menggunakan pustaka *scikit-learn*. Potongan kode program yang digunakan dalam proses *clustering* ditampilkan pada Gambar 5.

```
from sklearn import cluster

# KMeans dengan 2 cluster
k = 2
km = cluster.KMeans(n_clusters=2, init='random', max_iter=5000, tol=1e-5, random_state=1)
km.fit(X_scaled)
```

Gambar 5. Penerapan kode Python untuk proses *clustering* menggunakan algoritma K-Means

Metode K-Means dalam penelitian ini menghasilkan dua klaster wilayah berdasarkan kemiripan karakteristik akses terhadap fasilitas dasar dan tingkat pengeluaran rumah tangga. Data yang digunakan (X_scaled) merupakan data yang telah dinormalisasi menggunakan *Robust Scaler*. Hasil segmentasi ditampilkan pada Tabel 5, yang menunjukkan pembagian provinsi dalam masing-masing klaster.

Tabel 5. Hasil *clustering* menggunakan metode K-Means

klaster	Nama Provinsi
Klaster 1	Kalimantan Tengah, Sulawesi Tengah, Jawa Barat, Aceh, Kalimantan Utara, Bali, Jawa Timur, Sulawesi Utara, Daerah Istimewa Yogyakarta, Lampung, Kalimantan Timur, Kalimantan Barat, Sumatera Barat, Bengkulu, Nusa Tenggara Barat, Banten, Jawa Tengah, DKI Jakarta, Jambi, Gorontalo, Kepulauan Riau, Sumatera Selatan, Kalimantan Selatan, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Barat, Riau, Kepulauan Bangka Belitung, Sumatera Utara.
Klaster 2	Nusa Tenggara Timur, Maluku Utara, Maluku, Papua, Irian Jaya Barat (Papua Barat).

Setelah proses segmentasi dilakukan, langkah berikutnya adalah memahami lebih dalam karakteristik masing-masing klaster yang telah terbentuk. Analisis lebih lanjut dilakukan terhadap terhadap nilai akhir pusat klaster (*centroid*) dari masing-masing klaster untuk mengidentifikasi karakteristik wilayah dalam setiap kelompok. Tabel 6 menyajikan nilai *centroid* yang diperoleh dari rata-rata setiap variabel pada masing-masing klaster. Nilai *centroid* ini menjadi dasar dalam memahami perbedaan karakteristik antar provinsi berdasarkan fasilitas dasar dan pola pengeluaran rumah tangga.

Tabel 6. Nilai akhir pusat klaster (centroid) setiap variabel

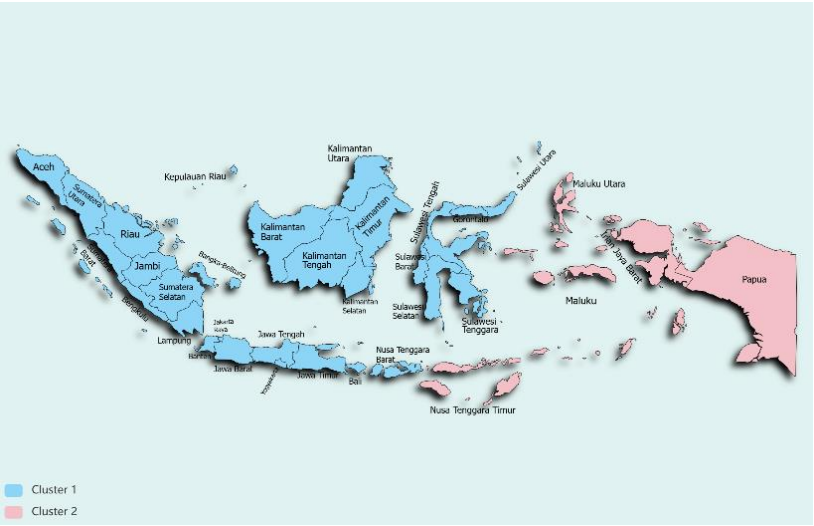
Variable	Klaster 1	Klaster 2
Kepemilikan Rumah	-0.221709	0.196941
Air minum layak	-0.086281	-0.542746
Sanitasi layak	0.171140	-1.759443
Listrik PLN	-0.186432	-4.268.487
Penggunaan Gas	-0.098821	-8.427.691
Rumah layak huni	0.030046	-0.851681
Pengeluaran listrik	0.495894	-0.464455
Pengeluaran untuk air	0.684524	-0.292709
Pengeluaran bahan bakar	-0.002504	1.982.252
Pengeluaran gas	0.294306	-2.933.136
Pengeluaran kebutuhan rumah tangga lainnya	0.319504	-0.174214

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 6, Nomor 4, Juni 2025
e-ISSN 2685-998X
DOI 10.30865/json.v6i4

Berdasarkan nilai *centroid* pada Tabel 6, klaster 1 mencakup provinsi-provinsi dengan nilai *centroid* yang lebih tinggi pada variabel *akses sanitasi layak, rumah layak huni, pengeluaran listrik, air, gas, dan kebutuhan rumah lainnya*. Ini menunjukkan bahwa wilayah-wilayah dalam klaster ini memiliki kondisi infrastruktur dasar yang relatif lebih baik dan tingginya pengeluaran rumah tangga yang menggambarkan kesejahteraan ekonomi yang lebih baik. Namun, nilai negatif pada variabel *kepemilikan rumah* menunjukkan bahwa proporsi kepemilikan rumah relatif lebih rendah yang mengindikasikan tingginya proporsi masyarakat yang tinggal di rumah kontrakan atau sewa, khususnya di kawasan perkotaan. Hal ini menunjukkan tantangan di sektor perumahan, terutama di wilayah perkotaan, dan perlunya kebijakan subsidi rumah pertama atau hunian terjangkau. Sebaliknya, klaster 2 menunjukkan nilai *centroid* yang jauh lebih rendah, bahkan sangat negatif, terutama pada variabel *akses sanitasi layak (-1.76), listrik (-4.27), dan penggunaan gas (-8.43)*. Hal ini mengindikasikan bahwa provinsi dalam klaster 2 menghadapi tantangan dalam pemenuhan infrastruktur dasar dan energi rumah tangga. Namun, nilai *rumah milik sendiri* lebih tinggi pada klaster ini. Selain itu, satu-satunya variabel yang menunjukkan nilai positif yang paling tinggi adalah *pengeluaran bahan bakar non-gas*, yang menunjukkan ketergantungan pada kayu bakar, minyak tanah, atau bahan bakar tradisional lainnya.

Kedua klaster menunjukkan nilai negatif pada variabel akses listrik dan penggunaan gas, menandakan masih perlunya peningkatan akses terhadap energi modern. Oleh karena itu, pembangunan infrastruktur energi terbarukan menjadi penting untuk mendukung pemerataan akses energi. Temuan ini sejalan dengan laporan pembangunan nasional yang menyoroti adanya kesenjangan antara Indonesia bagian barat dan timur, di mana wilayah barat, di mana kawasan barat terutama Pulau Jawa dan Sumatera, telah menjadi pusat pertumbuhan ekonomi dan infrastruktur , Sementara itu, wilayah timur masih menghadapi tantangan dalam hal akses layanan dasar [2]. Secara keseluruhan, hasil pengelompokan ini menunjukkan adanya kesenjangan pembangunan antarwilayah, yang sudah menjadi perhatian dalam agenda pemerataan pembangunan nasional. Oleh karena itu, pentingnya pendekatan berbasis wilayah (*place-based policy*) dalam perencanaan pembangunan, dengan menekankan pada peningkatan akses infrastruktur dasar di wilayah tertinggal guna memperkuat pemerataan pembangunan dan kesejahteraan di seluruh Indonesia.

Visualisasi hasil *clustering* ditampilkan dalam Gambar 6 yang menggambarkan peta spasial persebaran letak provinsi berdasarkan klaster, di mana provinsi dalam klaster 1 digambarkan dengan warna biru, sedangkan klaster 2 digambarkan dengan warna merah muda.



Gambar 6. Visualisasi Provinsi di Indonesia berdasarkan Hasil *Clustering* menggunakan K-Means

4. KESIMPULAN

Hasil penelitian ini mengindikasikan bahwa algoritma K-Means efektif dalam mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator akses terhadap fasilitas dasar dan tingkat pengeluaran rumah tangga, menghasilkan dua klaster dengan karakteristik yang berbeda. Klaster 1 terdiri dari 29 provinsi yang memiliki akses lebih baik terhadap fasilitas dasar, seperti rumah layak huni, air minum, dan sanitasi, serta menunjukkan pengeluaran rumah tangga yang lebih tinggi. Sebaliknya, klaster 2 mencakup 5 provinsi yang menghadapi tantangan lebih besar dalam hal infrastruktur dasar dan pengeluaran rumah tangga, dengan nilai *centroid* yang lebih rendah pada akses air minum, sanitasi, dan pengeluaran rumah tangga. Secara keseluruhan, perbedaan antara klaster 1 dan klaster 2 mencerminkan ketimpangan antara provinsi-provinsi di Indonesia, di mana klaster 1 menggambarkan provinsi yang lebih maju dalam hal infrastruktur dan kesejahteraan rumah tangga, sementara

klaster 2 menunjukkan provinsi yang memerlukan perhatian lebih dalam pengembangan fasilitas dasar dan infrastruktur. Hasil ini menandakan perlunya perhatian lebih dari pemerintah untuk meningkatkan infrastruktur dasar dan kesejahteraan masyarakat, terutama di provinsi-provinsi yang tergolong dalam klaster 2. Rencana penelitian selanjutnya dapat difokuskan pada analisis kebijakan pemerintah yang dapat mendukung pembangunan di provinsi-provinsi dalam klaster 2.

REFERENCES

[1] A. D. Laksono and R. D. Wulandari, "Urban – Rural Disparities of Facility-Based Childbirth in Indonesia," in *Proceedings of the 4th International Symposium on Health Research (ISHR 2019)*, Denpasar, Indonesia: Atlantis Press SARL, 2020, pp. 33–39. doi: 10.2991/ahsr.k.200215.007.

[2] Direktorat Jenderal Perimbangan Keuangan, "Laporan Perkembangan Ekonomi dan Fiskal Daerah Sinergi Pendanaan," 2021. [Online]. Available: https://djpk.kemenkeu.go.id/wp-content/uploads/2021/03/LPEFD-VI-Kinerja-Smart-City.pdf

[3] D. Setiawan, A. Nilogiri, and M. Dasuki, "Pengelompokan Provinsi Di Indonesia Berdasarkan Index Kesehatan Masyarakat Menggunakan Algoritma Partitioning Around Medoids (PAM) Dan Metode Davies Bouldin Index (DBI)," *J. Apl. Sist. Inf. dan Elektron.*, vol. 4, no. 1, pp. 10–16, 2022, [Online]. Available: http://jurnal.unmuhjember.ac.id/index.php/JASIE/article/view/20695%0Ahttp://jurnal.unmuhjember.ac.id/index.php/JASIE/article/download/20695/4648

[4] B. W. Otok, A. Suharsono, Purhadi, R. E. Standsyah, and H. Al Azies, "Partitional Clustering of Underdeveloped Area Infrastructure with Unsupervised Learning Approach: A Case Study in the Island of Java, Indonesia," *J. Reg. City Plan.*, vol. 33, no. 2, pp. 29–48, 2022, doi: 10.5614/jpwk.2022.33.2.3.

[5] N. B. Pratama, E. P. Purnomo, and A. Agustiyara, "Sustainable Development Goals (SDGs) dan Pengentasan Kemiskinan Di Daerah Istimewa Yogyakarta," *SOSIOHUMANIORA J. Ilm. Ilmu Sos. Dan Hum.*, vol. 6, no. 2, pp. 64–74, 2020, doi: 10.30738/sosio.v6i2.8045.

[6] P. Naskah *et al.*, *Indikator Perumahan dan Kesehatan Lingkungan*, vol. 9. Jakarta: Badan Pusat Statistik, 2023. [Online]. Available: https://www.bps.go.id/id/publication/2023/12/22/27008915741ff63ce2a2a054/housing-and-environmental-health-indicators-2023.html

[7] B. Wibawa, I. Fauzi, D. A. Novianti, N. Shabrina, A. D. Saputra, and S. A. Latief, "Development of Sustainable Infrastructure in Eastern Indonesia," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 832, no. 1, 2021, doi: 10.1088/1755-1315/832/1/012045.

[8] Rais, A. A. Dalimunthe, A. Fitrianto, B. Sartono, and S. D. Oktarina, "Regency Clusterization Based on Village Characteristics to Increase the Human Development Index (IPM) in Papua Province," *J. Ekon. Pembang.*, vol. 20, no. 02, pp. 153–168, 2022, doi: 10.22219/jep.v21i02.22911.

[9] T. R. Noviandy *et al.*, "Environmental and Economic Clustering of Indonesian Provinces: Insights from K-Means Analysis," *Leuser J. Environ. Stud.*, vol. 2, no. 1, pp. 41–51, 2024, doi: 10.60084/ljes.v2i1.181.

[10] S. and P. N. J. Russell, *Artificial Intelligence A Modern Appoarch Fourth Edition*, Fourth Edi. Pearson, 2022.

[11] P. Bintoro, Ratnasari, E. Wihardjo, I. P. Putri, and A. Asari, *Pengantar Machine Learning*, vol. 11, no. 1. PT Mafy Media Literasi Indonesia, 2024. [Online]. Available: http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETUNGAN_TERPUSAT_STRATEGI_MELESTARI

[12] R. Kesuma and A. Purwoto, "Pengelompokan Kabupaten/Kota Berdasarkan Indikator Rumah Layak Huni di Provinsi Jawa Barat Tahun 2020," in *Seminar Nasional Official Statistics*, 2022, pp. 995–1004. doi: 10.34123/semnasoffstat.v2022i1.1303.

[13] Muttaqin *et al.*, *Pengenalan Data Mining*. Yayasan Kita Menulis, 2023.

[14] I. Muthahharah, S. M. Meliyana, A. S. Ahmar, and A. Rahman, "K-Means Cluster for Grouping Regencies/Cities in South Sulawesi Province Based Human Development Index on the 2023," *ARRUS J. Soc. Sci. Humanit.*, vol. 4, no. 3, pp. 357–364, 2024, doi: 10.35877/soshum2587.

[15] M. . Rani Rotul Muhima, S.Si. *et al.*, *Kupas Tuntas Algoritma Clustering Konsep, Perhitungan Manual dan Program*. ANDI, 2021.

[16] A. M. Ikotun, M. S. Almutari, and A. E. Ezugwu, "K-means-based nature-inspired metaheuristic algorithms for automatic data clustering problems: Recent advances and future directions," *Appl. Sci.*, vol. 11, no. 23, pp. 1–61, 2021, doi: 10.3390/app112311246.

[17] C. A. Sugianto and T. P. O. R. Bokings, "K-Means Algorithm For Clustering Poverty Data in Bangka Belitung Island Province," *J. Comput. Networks, Archit. High-Performance Comput.*, vol. 3, no. 1, pp. 58–67, 2021, doi: 10.47709/cnahpc.v3i1.934.

[18] S. D. K. Wardani, A. S. Ariyanto, M. Umroh, and D. Rolliawati, "Perbandingan Hasil Metode Clustering K-Means, Db Scanner & Hierarchical Untuk Analisa Segmentasi Pasar," *JIKO (Jurnal Inform. dan Komputer)*, vol. 7, no. 2, pp. 191–201, 2023, doi: 10.26798/jiko.v7i2.796.

[19] S. Suwarsito, H. Mustafidah, T. Pinandita, and P. Purnomo, "Freshwater Fish Classification Based on Image Representation Using K-Nearest Neighbor Method," *JUITA J. Inform.*, vol. 10, no. 2, p. 183, 2022, doi: 10.30595/juita.v10i2.15471.

[20] M. Qori'atunnadyah and F. D. Rahmawati, "Pengelompokkan Kabupaten dan Kota Berdasarkan Kondisi Infrastruktur Jalan Menggunakan Hierarchical Clustering," *J. Informatics Dev.*, vol. 1, no. 1, pp. 1–5, 2022, doi: 10.30741/jid.v1i1.894.

[21] E. Widodo, P. Ermayani, L. N. Laila, and A. T. Madani, "Pengelompokkan Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Analisis Hierarchical Agglomerative Clustering," *Semin. Nas. Off. Stat.*, vol. 2021, no. 1,

pp. 557–566, 2021, doi: 10.34123/semnasoffstat.v2021i1.971.
[22] A. Novianti, I. M. Afnan, R. I. B. Utama, and E. Widodo, "Grouping of Districts Based on Poverty Factors in Papua Province Uses The K-Medoids Algorithm," *Enthusiastic Int. J. Appl. Stat. Data Sci.*, vol. 1, no. 2, pp. 94–102, 2020, doi: 10.20885/enthusiastic.vol1.iss2.art6.
[23] R. D. Faturahman and N. Hidayati, "Implementasi fuzzy c-means dalam pengelompokan tingkat kemiskinan pada kabupaten/kota di provinsi jawa tengah," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 1, pp. 137–149, 2025, doi: 10.29100/jipi.v10i1.5747.
[24] N. Puttanapong and S. Lim, "Predicting Household Expenditure Using Machine Learning Techniques: A Case of Cambodia," *Nakhara J. Environ. Des. Plan.*, vol. 23, no. 3, pp. 1–33, 2024, doi: 10.54028/NJ202423421.
[25] F. Andrianus and K. Alfatih, "Pengaruh Infrastruktur terhadap Kemiskinan: Analisis Data Panel 34 Provinsi di Indonesia," *J. Inform. Ekon. Bisnis*, vol. 5, no. 1, pp. 56–62, 2023, doi: 10.37034/infeb.v5i1.206.
[26] S. E. Wardani, S. Z. Harahap, and R. Muti'ah, "Implementation of the K-Means Method for Clustering Regency/City in North Sumatra based on Poverty Indicators," *Sink. J. dan Penelit. Tek. Inform.*, vol. 8, no. 3, pp. 1429–1442, 2024, doi: 10.33395/sinkron.v8i3.13720.
[27] K. D. R. Sianipar, S. W. Siahaan, and I. Gunawan, "Application of the K-Means algorithm in grouping households by province and ownership status of owned houses," *Sink. J. dan Penelit. Tek. Inform.*, vol. 5, no. 2, pp. 251–254, 2021, doi: 10.33395/sinkron.v5i2.10883.
[28] A. Septianingsih, "Analisis K-Means Clustering Pada Pemetaan Provinsi Indonesia Berdasarkan Indikator Rumah Layak Huni," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 3, no. 1, pp. 224–241, 2022, doi: 10.46306/lb.v3i1.116.
[29] A. A. A. Daniswara and I. K. D. Nuryana, "Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru," *J. Informatics Comput. Sci.*, vol. 05, no. 1, pp. 97–100, 2023.
[30] W. Nugraha, R. Sabaruddin, and S. Murni, "Teknik Scaling Menggunakan Robust Scaler Untuk Mengatasi Outlier Data Pada Model Prediksi Serangan Jantung," *Techno.Com*, vol. 23, no. 2, pp. 319–327, 2024, doi: 10.62411/tc.v23i2.10463.
[31] C. Paramita, F. A. Rafrastara, and C. Supriyanto, "Pemanfaatan Algoritma K-Means untuk Membuktikan Implementasi Undang-Undang Pelanggaran Hukum Korupsi di Pengadilan Negeri Banjarmasin," *J. Inform. J. Pengemb. IT*, vol. 8, no. 2, pp. 149–154, 2023, doi: 10.30591/jpit.v8i2.5216.
[32] A. B. Astuti, A. N. Guci, V. I. A. Alim, L. N. Azizah, M. K. Putri, and W. Ngabu, "Non Hierarchical K-Means Analysis To Clustering Priority Distribution of Fuel Subsidies in Indonesia," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 3, pp. 1663–1672, 2023, doi: 10.30598/barekengvol17iss3pp1663-1672.
[33] H. P. Kurniawan and L. Farhatuaini, "Identifikasi Pola Kepuasan Mahasiswa Terhadap Proses Pembelajaran Menggunakan Algoritma K-Means Clustering.," *J. Inform. J. Pengemb. IT*, vol. 9, no. 2, pp. 164–172, 2024, doi: 10.30591/jpit.v9i2.6740.
[34] D. Fuji Astri and M. Martanto, "Clustering Penduduk Miskin Menggunakan Algoritma K-Means Pada Wilayah Jawa Barat," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1548–1554, 2024, doi: 10.36040/jati.v8i2.9012.