

Pengelompokan Provinsi Indonesia Berdasarkan Rasio Penggunaan Gas Rumah Tangga Pada Tahun 2023 Menggunakan Hierarchical Clustering

Afdal Aditya Nastion¹, Pradita Eko Prasetyo Utomo*², Ulfa Khaira³, Akhiyar Waladi⁴

^{1,2,3,4}Sistem Informasi, Fakultas Sains Dan Teknologi, Universitas Jambi, Indonesia

e-mail: *¹afdaladitya0@gmail.com, ²pradita.eko@unja.ac.id, ³Ulfa.ilkom@gmail.com, ⁴Akhiyar.waladi@unja.ac.id

Abstract – The use of household gas energy is a crucial indicator of energy accessibility and distribution equity across Indonesian provinces. This study aims to cluster Indonesian provinces based on the ratio of household gas usage in 2023 using the Agglomerative Hierarchical Clustering method with an average linkage approach. Data was obtained from the Central Bureau of Statistics (BPS) and underwent preprocessing steps including data cleaning, normalization, and transformation. The number of clusters was determined using the elbow method, while cluster quality was evaluated using the Silhouette Coefficient. The analysis revealed that the optimal clustering consists of two clusters. Cluster 1, comprising most provinces, showed a high household gas usage ratio of 90.02%, while Cluster 2, which includes Maluku, East Nusa Tenggara, North Maluku, West Papua, and Papua, exhibited a low ratio of 2.3%. These findings highlight significant disparities in gas usage patterns, attributed to differences in infrastructure and geographical conditions. This study concludes that hierarchical clustering is effective in identifying energy usage patterns and provides valuable insights for policy formulation to ensure equitable energy distribution.

Keywords – Elbow Method, Average linkage, Silhouette Coefficient

Abstrak – Penggunaan energi gas rumah tangga merupakan indikator penting yang mencerminkan aksesibilitas dan pemerataan distribusi energi di berbagai provinsi Indonesia. Penelitian ini bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan rasio penggunaan gas rumah tangga tahun 2023 menggunakan metode Agglomerative Hierarchical Clustering dengan pendekatan average linkage. Data diperoleh dari Badan Pusat Statistik (BPS) dan melalui tahapan praproses seperti pembersihan, normalisasi, dan transformasi data. Jumlah kluster ditentukan menggunakan metode elbow, sedangkan kualitas kluster dievaluasi dengan Silhouette Coefficient. Analisis menunjukkan bahwa kluster optimal terdiri dari dua kluster. Kluster 1, yang mencakup sebagian besar provinsi, memiliki rata-rata rasio penggunaan gas rumah tangga sebesar 90,02%, sementara Kluster 2, yang meliputi Maluku, Nusa Tenggara Timur, Maluku Utara, Papua Barat, dan Papua, memiliki rasio yang rendah sebesar 2,3%. Hasil ini menunjukkan adanya kesenjangan signifikan dalam pola penggunaan gas rumah tangga akibat perbedaan infrastruktur dan kondisi geografis. Penelitian ini menyimpulkan bahwa metode klusterisasi hierarki efektif dalam mengidentifikasi pola penggunaan energi dan memberikan wawasan berharga untuk merumuskan kebijakan distribusi energi yang lebih merata.

Kata Kunci – Metode elbow, Average linkage, Silhouette Coefficient

I. PENDAHULUAN

Penggunaan energi gas rumah tangga adalah salah satu indikator penting dalam mencerminkan aksesibilitas dan distribusi energi di Indonesia. Gas rumah tangga telah menjadi sumber energi utama bagi sebagian besar masyarakat, terutama di wilayah perkotaan. Namun, distribusi dan pemanfaatan energi ini masih menghadapi berbagai tantangan, termasuk disparitas infrastruktur, aksesibilitas wilayah, serta kondisi sosial ekonomi yang berbeda-beda antar provinsi [1]. Ketimpangan ini dapat memengaruhi pencapaian pembangunan berkelanjutan, yang menjadi salah satu tujuan nasional.

Pemerintah Indonesia telah berupaya memperbaiki pemerataan distribusi energi, namun hasilnya masih belum optimal di beberapa wilayah. Oleh karena itu, diperlukan analisis yang mendalam untuk memahami pola konsumsi energi di setiap daerah. Pendekatan berbasis data, seperti klusterisasi, menjadi salah satu metode yang efektif untuk mengidentifikasi pola penggunaan energi [2]. Dengan metode ini, wilayah dapat dikelompokkan berdasarkan karakteristik serupa, sehingga dapat mendukung pengambilan keputusan kebijakan yang lebih tepat sasaran. Salah satu metode yang populer dalam klusterisasi adalah Agglomerative Hierarchical Clustering. Metode ini memiliki keunggulan dalam menyusun struktur hierarkis yang mampu menggambarkan hubungan antar data dengan jelas. Penelitian sebelumnya menunjukkan bahwa metode ini efektif dalam mengelompokkan data multivariat, termasuk data distribusi energi [3]. Evaluasi kualitas kluster menggunakan koefisien korelasi cophenetic dan Silhouette Coefficient telah terbukti andal dalam menentukan jumlah kluster optimal [4].

Dalam konteks penelitian ini, data rasio penggunaan gas rumah tangga tahun 2023 dari Badan Pusat Statistik (BPS) digunakan sebagai dasar untuk melakukan pengelompokan provinsi. Dengan menggunakan metode Agglomerative Hierarchical Clustering dan evaluasi berbasis data, diharapkan hasil klusterisasi dapat memberikan wawasan baru mengenai distribusi energi di Indonesia. Sebagai contoh, penelitian Pipit Mutiara Putri et al. berhasil menggunakan metode K-Means Clustering untuk menganalisis data konsumsi energi rumah tangga, yang memberikan pandangan berharga terhadap disparitas penggunaan energi di berbagai provinsi [5].

Pendekatan yang dilakukan dalam penelitian ini juga memiliki kebaruan dalam menggunakan data terbaru tahun 2023, yang memungkinkan pengelompokan lebih relevan terhadap kondisi saat ini. Penelitian ini bertujuan untuk menghasilkan pengelompokan provinsi yang dapat menjadi landasan bagi pemerintah dalam menyusun kebijakan distribusi energi yang lebih merata. Selain itu, penelitian ini juga diharapkan dapat memberikan kontribusi ilmiah pada pengembangan metode klusterisasi untuk aplikasi dalam analisis distribusi energi [6][7].

Dengan adanya penelitian ini, diharapkan dapat dihasilkan rekomendasi kebijakan yang tidak hanya meningkatkan efisiensi distribusi energi gas rumah tangga, tetapi juga mendukung tujuan pembangunan berkelanjutan di Indonesia [8].

II. PENELITIAN YANG TERKAIT

Kajian terkait pengelompokan provinsi di Indonesia berdasarkan rasio penggunaan gas rumah tangga melalui metode clustering menjadi sangat penting untuk memahami pola konsumsi energi masyarakat. Salah satu penelitian yang relevan dilakukan oleh Pipit Mutiara Putri dan tim, yang dipublikasikan dalam prosiding Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) pada tahun 2020. Penelitian tersebut menggunakan metode K-Means Clustering untuk menganalisis data rasio penggunaan gas rumah tangga di Indonesia pada periode 2015-2018 [5]. Tujuan utama dari penelitian ini adalah untuk menggambarkan pola serta variasi tingkat adopsi gas rumah tangga di berbagai provinsi.

Hasil penelitian menunjukkan bahwa provinsi di Indonesia dapat dikelompokkan berdasarkan pola konsumsi gas rumah tangga menjadi beberapa kategori. Klasifikasi tersebut mencerminkan adanya perbedaan dalam tingkat penggunaan, di mana beberapa provinsi menunjukkan penggunaan yang tinggi, sementara lainnya masih rendah. Beberapa faktor yang memengaruhi variasi ini meliputi ketersediaan infrastruktur distribusi gas, aksesibilitas wilayah, serta kebijakan yang diterapkan oleh pemerintah daerah [5].

Berdasarkan analisis, provinsi yang memiliki tingkat penggunaan gas rumah tangga tertinggi umumnya adalah daerah perkotaan dengan akses infrastruktur yang memadai. Sebaliknya, provinsi dengan tingkat adopsi rendah biasanya berada di wilayah pedesaan atau terpencil. Penelitian ini berkontribusi signifikan terhadap penyusunan kebijakan energi yang bertujuan meningkatkan pemerataan akses gas rumah tangga [5].

Penelitian yang terkait menjadi acuan untuk dilakukannya penelitian ini, metode hierarchical clustering dapat diterapkan untuk memperoleh wawasan yang lebih mendalam mengenai struktur hierarkis pengelompokan provinsi. Dengan menggunakan data terbaru, seperti rasio penggunaan gas rumah tangga tahun 2023, hasil analisis dapat mendukung perencanaan kebijakan distribusi gas yang lebih merata dan tepat sasaran.

III. METODE PENELITIAN

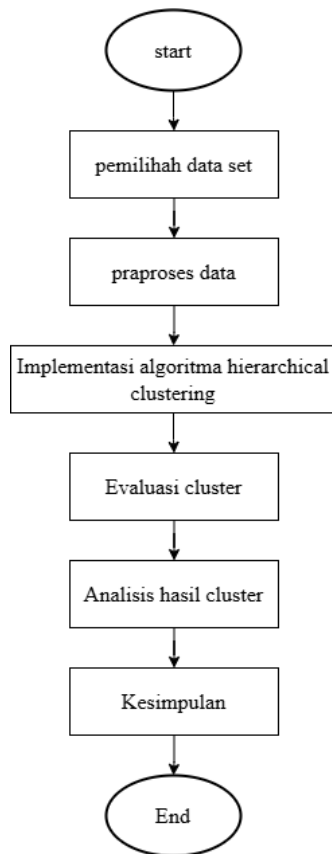
A. Pengumpulan Data

a. Studi Dokumen

Data yang digunakan dalam penelitian ini adalah data rasio penggunaan gas rumah tangga per provinsi di Indonesia dari Badan Pusat Statistik (BPS) tahun 2023. Data ini mencakup informasi kuantitatif tentang proporsi penggunaan gas sebagai sumber energi rumah tangga di berbagai provinsi.

B. Metode Penelitian

Metode penelitian adalah kerangka kerja atau pendekatan ilmiah yang digunakan untuk mengumpulkan, menganalisis, dan menafsirkan data guna menjawab pertanyaan penelitian. Pada penelitian pengelompokan provinsi di Indonesia berdasarkan rasio penggunaan gas rumah tangga menggunakan hierarchical clustering ini ada beberapa tahapan-tahapan yang dilakukan yaitu pemilihan data set, praproses data, implementasi algoritma hierarchical clustering, evaluasi cluster, dan analisis hasil cluster. Untuk lebih jelasnya dapat dilihat dari gambar 1:



Gambar 1 tahapan penelitian

a. Pemilihan data set

Pemilihan dataset adalah proses menentukan kumpulan data yang relevan, berkualitas, dapat diakses, dan sesuai dengan tujuan penelitian untuk memastikan validitas, keandalan, serta keakuratan hasil analisis [9]. Data yang digunakan dalam penelitian ini adalah data rasio penggunaan gas rumah tangga per provinsi di Indonesia dari Badan Pusat Statistik (BPS) tahun 2023. Data yang digunakan dapat dilihat pada tabel 1.

TABLE 1
DATA SET RASIO PENGGUNAAN GAS RUMAH TANGGA

Provinsi	Rasio Penggunaan Gas Rumah Tangga 2023
ACEH	94.53
SUMATERA UTARA	92.62
SUMATERA BARAT	87.11
RIAU	95.55
JAMBI	92.53
SUMATERA SELATAN	96.37
BENGGULU	95.63
LAMPUNG	92.41
....
PAPUA	3.03
INDONESIA	86.91

b. Praproses Data

1. Pembersihan Data (Data Cleaning)

Data Cleaning adalah proses mengidentifikasi dan menghapus data duplikat atau missing value yang dapat menghasilkan hasil yang tidak akurat [10].

2. Transformasi Data
Transformasi data adalah proses Standarisasi dan Normalisasi untuk Menyesuaikan skala dan bentuk distribusi data agar lebih konsisten [10]. Adapun metode yang digunakan yaitu metode min-max. Metode MinMax Normalization adalah metode normalisasi yang merubah rentang nilai data menjadi antara 0 dan 1 [11].
- c. Implementasi Agglomerative Hierarchical Clustering
 1. Pemilihan Metode Klasterisasi
Pada tahap ini, hasil klasterisasi dari setiap metode linkage dibandingkan dengan menggunakan koefisien korelasi cophenetic untuk mengevaluasi kualitas klaster yang terbentuk. Koefisien ini mengukur seberapa baik struktur klaster yang dihasilkan oleh metode tersebut mencerminkan jarak atau kesamaan dalam data aslinya. Semakin mendekati nilai koefisien ini ke 1, semakin baik struktur klaster yang dihasilkan dalam merepresentasikan data secara akurat, sehingga metode linkage yang memberikan nilai tertinggi dianggap paling sesuai untuk dataset yang digunakan [12].
 2. Penerapan metode elbow
Metode elbow adalah teknik yang digunakan untuk menentukan jumlah cluster optimal dalam analisis clustering dengan mengidentifikasi titik "siku" (elbow) pada grafik yang menunjukkan hubungan antara jumlah cluster dan nilai total sum of squared errors (SSE), di mana penurunan SSE mulai melambat secara signifikan setelah titik tersebut [13].
 3. Penerapan Metode Agglomerative Hierarchical Clustering
mengimplementasikan metode Agglomerative Hierarchical Clustering kedalam Rstudio.
- d. Evaluasi Hasil Klasterisasi
Untuk menilai kualitas dan keoptimalan pembagian klaster, digunakan nilai rata-rata *Silhouette Coefficient* sebagai metrik evaluasi. Koefisien ini mengukur sejauh mana data dalam setiap klaster dikelompokkan secara homogen serta seberapa jelas perbedaan antar klaster yang terbentuk. Nilai koefisien yang mendekati 1 menunjukkan bahwa data berada pada klaster yang tepat, dengan jarak yang signifikan dari klaster lainnya. Sebaliknya, nilai mendekati 0 atau negatif menunjukkan bahwa data tersebut berada di perbatasan antar klaster atau mungkin salah pengelompokan, sehingga pembagian klaster memerlukan penyesuaian lebih lanjut [14].
- e. Analisis
Pada tahap ini, dilakukan analisis terhadap hasil klasterisasi untuk menafsirkan pola-pola yang terbentuk. Setiap klaster dianalisis untuk memahami karakteristik konsumsi gas rumah tangga di dalamnya, serta perbedaan antar klaster. Analisis ini bertujuan untuk mengidentifikasi wilayah-wilayah yang membutuhkan intervensi kebijakan lebih lanjut dalam penyediaan energi.

IV. HASIL DAN PEMBAHASAN

Penelitian analisis klaster ini mencakup beberapa tahap, yaitu praproses data, visualisasi awal, dan analisis data menggunakan metode klustering. Pada tahap praproses dan penyajian data, grafik, tabel serta analisis klaster dilakukan menggunakan software Rstudio.

A. Praproses data

Pada tahap ini data rasio penggunaan gas rumah tangga pada tahun 2023 akan melewati tahap praproses data. Data terdiri dari 35 data termasuk data rasio satu indonesia. Dari total 35 baris data, kemungkinan masih terdapat baris yang memiliki nilai atribut yang hilang (*missing value*), data yang duplikat, atau data yang tidak konsisten. Dalam penelitian ini, tahap praproses data meliputi tahapan pembersihan (*cleaning*) dan transformasi. Pada tahap pembersihan, data diperiksa dan dihilangkan dari nilai yang hilang serta duplikasi, sehingga menghasilkan data yang bersih seperti terlihat pada gambar 2:

Provinsi	Rasio_2023
ACEH	94.53
SUMATERA UTARA	92.62
SUMATERA BARAT	87.11
RIAU	95.55
JAMBI	92.53
SUMATERA SELATAN	96.37
BENGKULU	95.63
LAMPUNG	92.41
KEP. BANGKA BELITUNG	96.74
KEP. RIAU	90.87
DKI JAKARTA	96.81
JAWA BARAT	92.74
JAWA TENGAH	88.89
DI YOGYAKARTA	83.55
JAWA TIMUR	87.66
BANTEN	91.47
BALI	85.61
NUSA TENGGARA BARAT	89.85
NUSA TENGGARA TIMUR	11.43
KALIMANTAN BARAT	95.84
KALIMANTAN TENGAH	92.7
KALIMANTAN SELATAN	91.47
KALIMANTAN TIMUR	97.74
KALIMANTAN UTARA	87.13
SULAWESI UTARA	84.22
SULAWESI TENGAH	72.57
SULAWESI SELATAN	93.52
SULAWESI TENGGARA	67.87
GORONTALO	96.14
SULAWESI BARAT	83.57
MALUKU	11.13
MALUKU UTARA	11.67
PAPUA BARAT	14.24
PAPUA	3.03
INDONESIA	86.91

Gambar 1. dataset setelah cleaning

Dapat dilihat dari gambar 3 jumlah data tidak berubah dan tetap sebanyak 35 record, ini menunjukkan bahwa tidak ada *missing value* atau duplikasi dalam dataset ini. Maka tahap berikutnya adalah transformasi data dengan teknik normalisasi, di mana data numerik dinormalisasi menggunakan metode min-max untuk menghasilkan nilai dalam rentang 0 hingga 1.

	Rasio_2023
ACEH	0.966773626
SUMATERA UTARA	0.947003416
SUMATERA BARAT	0.889969982
RIAU	0.977331539
JAMBI	0.946071835
SUMATERA SELATAN	0.985819273
BENGKULU	0.978159611
LAMPUNG	0.944829728
KEP. BANGKA BELITUNG	0.989649105
KEP. RIAU	0.928889349
DKI JAKARTA	0.990373667
JAWA BARAT	0.948245523
JAWA TENGAH	0.908394576
DI YOGYAKARTA	0.853120795
JAWA TIMUR	0.895662975
BANTEN	0.935099886
BALI	0.874443639
NUSA TENGGARA BARAT	0.918331436
NUSA TENGGARA TIMUR	0.003105269
KALIMANTAN BARAT	0.980333299
KALIMANTAN TENGAH	0.947831487
KALIMANTAN SELATAN	0.935099886
KALIMANTAN TIMUR	1.000000000
KALIMANTAN UTARA	0.890177000
SULAWESI UTARA	0.860055895
SULAWESI TENGAH	0.739467964
SULAWESI SELATAN	0.956319222
SULAWESI TENGGARA	0.690818756
GORONTALO	0.983438567
SULAWESI BARAT	0.853327813
MALUKU	0.000000000
MALUKU UTARA	0.005589483
PAPUA BARAT	0.032191285
PAPUA	0.019666701
INDONESIA	0.887899803

Gambar 2. data setelah di normalisasikan

B. Nilai Koefisien Korelasi Cophenetic

Sebelum lanjut ke *agglomeratif hierarchical clustering* perlu untuk membandingkan performa hasil analisis klastering menggunakan metode *Agglomerative Hierarchical Clustering* dengan empat pendekatan yaitu *average linkage*,

complete linkage, single linkage, dan ward. Keefektifan hasil klustering dari masing-masing metode diukur menggunakan nilai koefisien korelasi *cophenetic*. Semakin mendekati nilai koefisien ini ke 1, semakin baik performa kluster yang dihasilkan. Adapun nilai koefisien korelasi *cophenetic* untuk setiap metode kluster yang diuji adalah sebagai berikut.

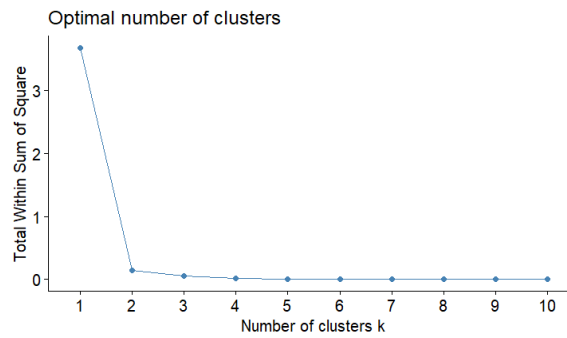
Method	Cophenetic_Correlation
1 single	0.9913961
2 complete	0.9918304
3 average	0.9933313
4 ward.D	0.9876968

Gambar 3. nilai koefisien korelasi *cophenetic*

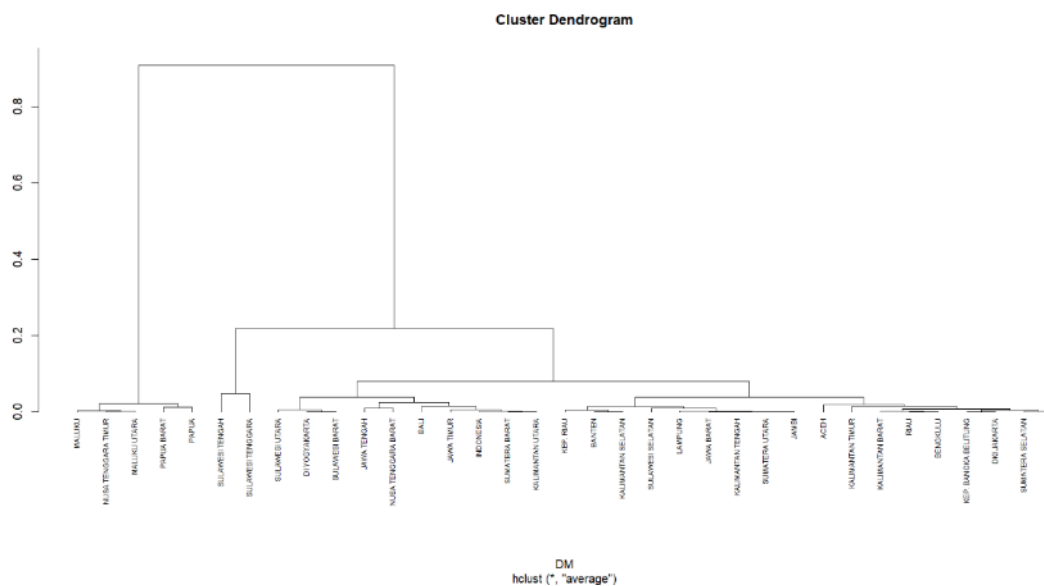
Dapat dilihat dari gambar 4 menunjukkan nilai koefisien korelasi *Cophenetic* pada masing-masing metode *agglomeratif hierarchical clustering*. Berdasarkan gambar 4 yang menunjukkan bahwa metode average linkage memiliki nilai koefisien korelasi *cophenetic* yang paling mendekati 1.

C. Penerapan Agglomerative Hierarchical Clustering

Sebelum melakukan proses klustering, maka perlu untuk menghitung matriks jarak antar data dengan menggunakan metode *Euclidean Distance*. Pada metode *Agglomerative*, klusterisasi dimulai dengan setiap data sebagai kluster terpisah (N kluster), yang kemudian digabungkan bertahap menjadi satu kluster tunggal dengan N adalah nilai dari jumlah data, maka jumlah awal cluster pada *agglomerative hierarchical clustering* adalah sebanyak jumlah data yaitu 35. Adapun pemilihan jumlah cluster akan dilakukan menggunakan metode elbow dengan berdasarkan pengamatan terhadap perubahan **Within-Cluster Sum of Squares (WSS)**, yaitu total jarak kuadrat antara titik data dan pusat klusternya. Hasil metode elbow dan klusterisasi divisualisasikan dalam dendrogram seperti terlihat pada Gambar 5 dan 6.

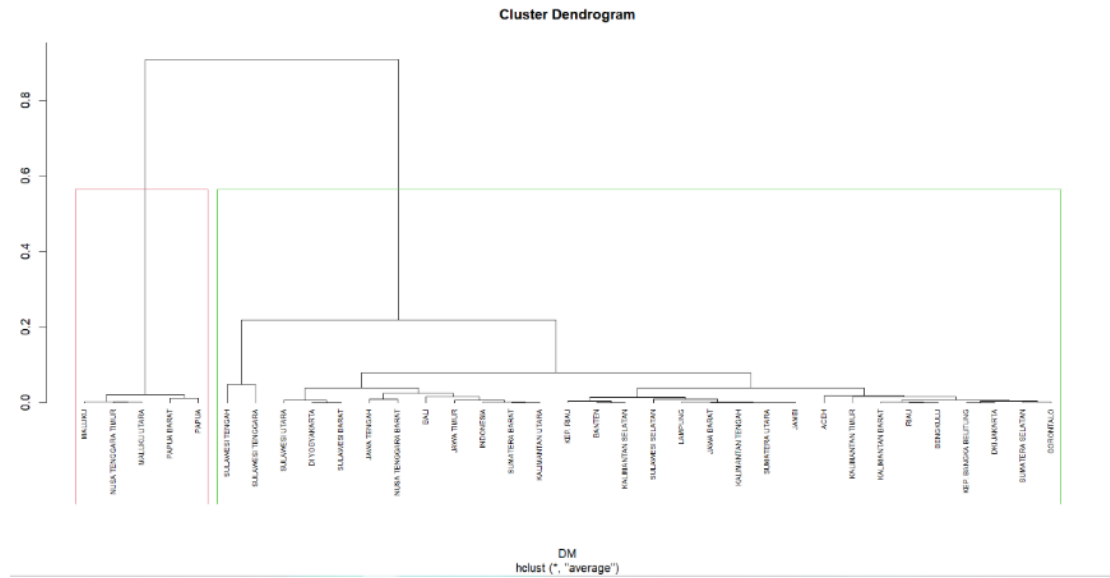


Gambar 4 metode elbow untuk memilih jumlah cluster



Gambar 5. visualisasi kluster hirarki dengan dendrogram`

Sesuai dengan nilai dari koefisien korelasi cophenetic maka metode kluster yang digunakan adalah average linkage dengan pemilihan jumlah kluster yang diambil dari metode elbow adalah 2. Hasil dari pemotongan pohon hirarki dapat dilihat pada gambar 7.



Gambar 6. klastering hirarki dengan k=2

Dari gambar 6 dapat dilihat setelah memotong pohon hirarki menjadi 2 kluster, maka terdapat 2 kelompok data yang terpisah dengan data kluster 1 sebanyak 30 data dan kluster 2 sebanyak 5 data seperti yang terdapat pada gambar 8.

sub_grp	
1	2
30	5

Gambar 7. table pemotongan pohon hirarki

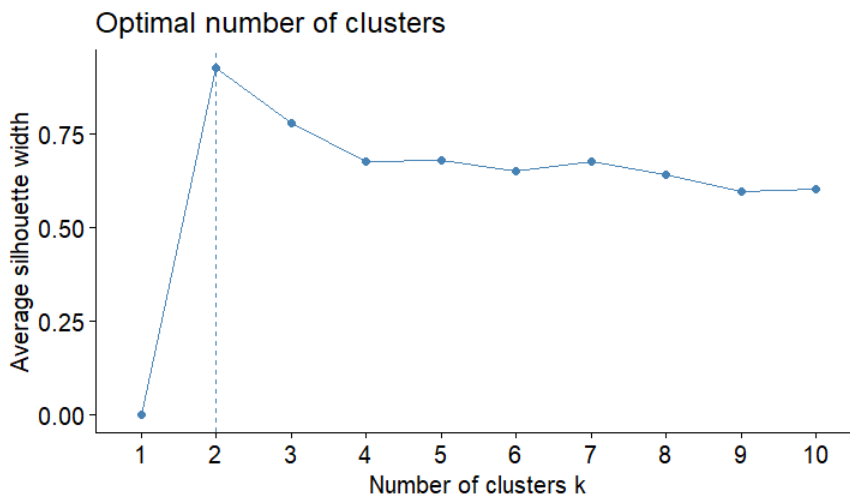
D. Evaluasi Hasil Cluster

Untuk evaluasi kluster dilakukan dengan menganalisis rata-rata nilai *silhouette coefficient* untuk berbagai jumlah kluster yang dicoba. Pengujian melibatkan variasi jumlah kluster mulai dari 1 hingga 10, nilai *silhouette coefficient* yang diperoleh untuk setiap jumlah kluster tercantum dalam gambar 9.

clusters	y
1	1 0.000000
2	2 0.9245789
3	3 0.7776386
4	4 0.6746746
5	5 0.6775446
6	6 0.6481349
7	7 0.6761941
8	8 0.6393516
9	9 0.5952596
10	10 0.6009398

Gambar 8. rata-rata nilai *silhouette coefficient*

Adapun hasil pengujian *silhouette coefficient* dalam bentuk grafik dapat dilihat pada gambar 10.



Gambar 9. rata-rata nilai *silhouette coefficient* dengan grafik

Berdasarkan hasil pengujian *silhouette coefficient* yang terlihat pada gambar 9, dengan mengesampingkan jumlah kluster 1, maka jumlah kluster yang memiliki nilai tertinggi adalah 2 dan jumlah kluster dengan nilai terendah adalah 9. Sehingga jumlah pengelompokan yang paling tepat dalam menganalisis penggunaan gas rumah tangga adalah 2 kelompok (kluster). Jumlah kluster 2 juga menunjukkan bahwa tingkat kohesi yang lebih tinggi dan pemisahan yang lebih jelas antar kluster [15].

E. Analisis

Pengujian terhadap nilai Silhouette Coefficient menunjukkan bahwa pembagian data menjadi dua kluster merupakan pilihan yang paling optimal dalam penelitian ini. Nilai Silhouette Coefficient yang mendekati 1 mengindikasikan kualitas kluster yang baik, di mana masing-masing kelompok memiliki tingkat keseragaman yang tinggi dalam kluster tersebut serta perbedaan yang jelas dengan kluster lainnya [16]. Dalam hal ini, hasil analisis mendukung bahwa pembagian menjadi dua kluster berhasil menggambarkan pola penggunaan gas rumah tangga di Indonesia pada tahun 2023.

Cluster	Rasio_2023
1	90.02067
2	2.30000

Gambar 10. rata-rata rasio penggunaan gas rumah tangga perkluster

Hasil klusterisasi menunjukkan adanya dua kelompok provinsi dengan karakteristik yang berbeda terkait penggunaan gas rumah tangga. Kluster 1 mencakup sebagian besar provinsi di Indonesia, dengan rata-rata rumah tangga yang menggunakan gas sebagai bahan bakar utama mencapai 90,02%. Kelompok ini cenderung didominasi oleh wilayah yang memiliki infrastruktur energi yang sudah mapan, terutama di daerah perkotaan dan semi-perkotaan. Ketersediaan distribusi energi yang memadai di wilayah-wilayah tersebut tampaknya menjadi salah satu faktor utama tingginya rasio penggunaan gas rumah tangga. Selain itu, faktor lain seperti urbanisasi yang relatif tinggi dan adanya kebijakan yang mendukung pemanfaatan energi ramah lingkungan kemungkinan turut mendorong tingginya tingkat adopsi gas sebagai bahan bakar.

Sebaliknya, Kluster 2 terdiri dari lima provinsi, yaitu Maluku, Nusa Tenggara Timur (NTT), Maluku Utara, Papua Barat, dan Papua. Kluster ini menunjukkan rata-rata penggunaan gas rumah tangga sebesar 2,3%, yang mencerminkan tingkat adopsi gas yang sangat rendah. Wilayah-wilayah dalam kluster ini umumnya menghadapi kendala geografis yang signifikan, seperti lokasi yang terpencil, sehingga akses terhadap distribusi energi gas menjadi terbatas. Selain itu, masyarakat di provinsi-provinsi ini masih cenderung menggunakan bahan bakar tradisional, seperti kayu bakar atau minyak tanah, untuk memenuhi kebutuhan energi mereka. Kekurangan infrastruktur energi yang memadai dan minimnya kebijakan pemerintah yang secara khusus mendukung penyediaan gas rumah tangga menjadi hambatan utama dalam peningkatan adopsi gas di wilayah-wilayah ini.

Temuan ini memberikan wawasan penting bagi pengambil kebijakan untuk merancang strategi distribusi energi yang lebih adil dan merata. Wilayah-wilayah dalam Kluster 2 memerlukan perhatian khusus melalui pembangunan infrastruktur energi yang lebih baik, penyaluran subsidi yang tepat sasaran, serta edukasi masyarakat tentang manfaat penggunaan gas rumah tangga. Dengan langkah-langkah tersebut, ketimpangan akses energi antara wilayah barat dan timur

Indonesia dapat diminimalkan, sehingga mendukung tercapainya pembangunan energi yang berkelanjutan. Di sisi lain, Klaster 1 dapat dijadikan contoh keberhasilan yang dapat diterapkan di wilayah lain, misalnya dengan memperluas jaringan distribusi gas atau mendorong konversi energi berbasis gas di daerah yang memiliki tingkat adopsi rendah.

Analisis ini memberikan pemahaman menyeluruh mengenai pola distribusi penggunaan gas rumah tangga di Indonesia. Selain itu, temuan ini juga menawarkan rekomendasi strategis untuk intervensi kebijakan yang lebih efektif dan berbasis data.

V. KESIMPULAN

Berdasarkan penelitian yang dilakukan, Penelitian ini telah berhasil melakukan pengelompokan provinsi-provinsi di Indonesia berdasarkan rasio penggunaan gas rumah tangga pada tahun 2023 dengan menggunakan metode Agglomerative Hierarchical Clustering dan pendekatan average linkage. Analisis yang dilakukan menunjukkan bahwa dua klaster merupakan jumlah yang optimal. Klaster 1 memiliki rata-rata rasio penggunaan gas rumah tangga sebesar 90,02%, yang mencakup mayoritas provinsi dengan infrastruktur energi yang memadai. Sementara itu, Klaster 2 menunjukkan rata-rata rasio penggunaan gas rumah tangga sebesar 2,3% dan terdiri dari provinsi-provinsi yang menghadapi tantangan geografis serta keterbatasan infrastruktur. Penilaian menggunakan Silhouette Coefficient mengonfirmasi bahwa klasterisasi yang dihasilkan memiliki kualitas yang baik, dengan kohesi internal yang tinggi dan pemisahan antar-klaster yang jelas. Rendahnya tingkat penggunaan gas di Klaster 2 dapat dikaitkan dengan kendala distribusi energi, kurangnya kebijakan pendukung, dan preferensi masyarakat terhadap bahan bakar tradisional.

Hasil penelitian ini memberikan wawasan yang penting untuk merancang kebijakan distribusi energi yang lebih merata di Indonesia. Provinsi dalam Klaster 2 membutuhkan perhatian khusus, seperti pengembangan infrastruktur distribusi gas, pemberian subsidi yang terarah, serta edukasi mengenai manfaat penggunaan gas rumah tangga. Di sisi lain, pendekatan yang berhasil diterapkan di Klaster 1 dapat dijadikan model untuk meningkatkan adopsi gas di wilayah lain. Penelitian ini merekomendasikan untuk memanfaatkan data yang lebih luas dan terkini serta mengintegrasikan metode analisis tambahan dalam penelitian di masa depan. Dengan demikian, hasil yang diperoleh dapat mendukung pengembangan kebijakan energi yang lebih efektif, adil, dan berkelanjutan.

UCAPAN TERIMA KASIH

Dengan rasa syukur yang mendalam, penulis mengucapkan terima kasih kepada Tuhan Yang Maha Esa atas segala rahmat, pertolongan, dan karunia-Nya yang tak terhingga. Segala puji hanya bagi-Nya yang telah memberikan petunjuk, kekuatan, dan kemudahan selama perjalanan dalam penyusunan penelitian ini. Penulis juga ingin menyampaikan rasa terima kasih kepada para dosen yang dengan penuh dedikasi telah memberikan bimbingan, ilmu, motivasi, dan arahan yang sangat berharga sepanjang proses penelitian ini. Tanpa kontribusi mereka, penelitian ini tidak akan mencapai hasil yang memuaskan.

DAFTAR PUSTAKA

- [1]. Meilinda Suriani Harefa, Syukri Hidayat, Grace Mercy Epsilon Hia, Sabda Yanti Pasaribu, Natasya Kaila Putri, and Mhd Asa Arzaki Muis, "Analisis Pemanfaatan Gas Alam sebagai Pengganti Gas LPG", sosial, vol. 2, no. 4, pp. 179–182, Dec. 2024.
- [2]. Regina Citra Kurnia Pangestu and Anak Agung Ketut Ayuningsasi, "Pengaruh Konsumsi Energi Sektor Industri, Rumah Tangga, dan Transportasi terhadap Emisi Karbon di Indonesia", Inisiatif, vol. 3, no. 4, pp. 297–311, Aug. 2024.
- [3]. Sasmita and S. Muntari, "Penerapan Algoritma K-Means Clustering pada Data Keluhan Pelanggan PT. PLN Persero Kota Pagar Alam," Jurnal Ilmiah Teknosains, vol. 9, no. 1, pp. 8–12, May 2023.
- [4]. M. Farid, F. Insani, S. Agustian, and L. Afriyanti, "Clustering Electricity Distribution Data Using Density-Based Spatial Clustering of Applications With Noise (DBSCAN) Algorithm," MALCOM: Indonesian Journal of Machine Learning and Computer Science, vol. 4, no. 3, pp. 1024–1033, July 2024, doi: 10.57152/malcom.v4i2.1426.
- [5]. Putri, P. M., et al. (2020). "Pengelompokan Data Rasio Penggunaan Gas Rumah Tangga Berdasarkan Provinsi di Indonesia Menggunakan Metode K-Means Clustering". *Prosiding Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*.
- [6]. M. Nazer and H. Handra, "Analisis Konsumsi Energi Rumah Tangga Perkotaan di Indonesia: Periode Tahun 2008 dan 2011," Jurnal Ekonomi dan Pembangunan Indonesia, vol. 16, no. 2, pp. 141–153, Jan. 2016, doi: 10.21002/jepi.v16i2.04.
- [7]. C. Astria, A. P. Windarto, A. Wanto, dan E. Irawan, "Metode K-Means pada Pengelompokan Wilayah Pendistribusian Listrik," Seminar Nasional Sains & Teknologi Informasi (SENSASI), vol. 2019, pp. 306–312, Juli 2019. [Online]. Tersedia: <http://prosiding.seminar-id.com/index.php/sensasi/issue/archive>. ISBN: 978-602-52720-2-8.
- [8]. T. Khotimah, A. Syukur, dan M. A. Soeleman, "Clustering Trafo Distribusi Menggunakan Algoritma Self-Organizing Map," Jurnal SIMETRIS, vol. 8, no. 1, pp. 15–20, Apr. 2017. DOI: 10.24176/simet.v8i1.808. [Online]. Tersedia: <https://www.researchgate.net/publication/316589172>.
- [9]. G. Hermawan, "Memahami Peran Dataset dalam Penelitian Kecerdasan Buatan: Kualitas, Aksesibilitas, dan Tantangan," Preprint, Oct. 2024. [Online]. Available: <https://www.researchgate.net/publication/384863552>.
- [10]. Rahayu, P., Sudipa, I. G. I., Suryani, Surachman, A., Ridwan, A., Darmawiguna, I. G. M., Sutoyo, M., Slamet, I., Harlina, S., & May Sanjaya, I. M., "Buku Ajar Data Mining", Vol. 1, Issue January, 2024.
- [11]. Permana, I., & Salisah, F. N. S., "Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation", *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, 2(1), 67–72, 2022.
- [12]. T. Apriliana dan E. Widodo, "Analisis Cluster Hierarki untuk Pengelompokan Provinsi di Indonesia berdasarkan Jumlah Base Transceiver Station dan Kekuatan Sinyal," KONSTELASI: Konvergensi Teknologi dan Sistem Informasi, vol. 3, no. 2, pp. 286–296, Desember 2023. [Online]. Tersedia: <https://www.uui.ac.id>.

- [13]. N. A. Maori and Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Jurnal SIMETRIS*, vol. 14, no. 2, pp. 277–287, Nov. 2023.
- [14]. D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Jurnal Matrix*, vol. 9, no. 3, pp. 102–109, Nov. 2019, doi: 10.31940/matrix.v9i3.1662.
- [15]. Syafiyah, U., Asrafi, I., Wicaksono, B., Puspitasari, D. P., & Sirait, M., "Analisis Perbandingan Metode Cluster Data Indikator Ketenagakerjaan di Jabar Tahun 2020", *Seminar Nasional Official Statistics, 2022*, 803–812, 2020
- [16]. Septianingsih, A., "Pemetaan Kabupaten Kota Di Provinsi Jawa Timur Berdasarkan Tingkat Kasus Penyakit Menggunakan Pendekatan Agglomeratif Hierarchical Clustering", *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 3(2), 367–386, 2022