

Penerapan *K-Means Clustering* Untuk Mengelompokkan Tingkat Konsumsi Listrik Menurut Provinsi di Indonesia

DOI: <http://dx.doi.org/10.35889/progresif.v21i2.2876>

Creative Commons License 4.0 (CC BY – NC)



Yenni Lutfhi Berliana¹, Irwansyah^{2*}

Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta Timur, Indonesia

*e-mail Corresponding Author: irwansyah@uhamka.ac.id

Abstract

Electricity consumption in Indonesia continues to increase every year, but its distribution is not evenly distributed across provinces. Therefore, an analysis is needed to group regions based on electricity consumption patterns to help plan more efficient distribution. This study aims to group regions in Indonesia based on their electricity usage patterns, considering customer types such as households, industry, business, social, government buildings, and street lighting. The clustering process is performed using the k-means clustering method. The data used is official PLN data from 2019 to 2024. The analysis process is performed using RapidMiner, with steps of data preprocessing, application of K-Means, and evaluation of the results using the Davies-Bouldin Index (DBI) method. The results of the study show that the provinces in Indonesia are divided into two clusters, namely Cluster 0 with 29 provinces with low-medium consumption and Cluster 1 with 5 provinces with high consumption, especially in the Java Island region. The DBI value of 0.507 indicates that the resulting clustering is quite optimal. These results are in line with the PLN report and should support more targeted infrastructure planning and power distribution policies.

Keywords: Data Mining; K-Means Clustering; Electricity Consumption

Abstrak

Konsumsi energi listrik di Indonesia terus meningkat setiap tahunnya, namun distribusinya tidak merata di seluruh provinsi. Oleh karena itu, diperlukan analisis untuk mengelompokkan wilayah berdasarkan pola konsumsi energi listrik guna membantu perencanaan distribusi yang lebih efisien. Penelitian ini bertujuan untuk mengelompokkan wilayah-wilayah di Indonesia berdasarkan pola penggunaan listriknya, dengan mempertimbangkan jenis pelanggan seperti rumah tangga, industri, bisnis, sosial, gedung pemerintahan, dan penerangan jalan umum. Proses pengelompokan dilakukan dengan menggunakan metode *k-means clustering*. Data yang digunakan merupakan data resmi dari PLN tahun 2019 hingga 2024. Proses analisis dilakukan menggunakan RapidMiner, dengan tahapan pre-processing data, penerapan *K-Means*, dan evaluasi hasil menggunakan metode *Davies-Bouldin Index* (DBI). Hasil penelitian menunjukkan bahwa provinsi di Indonesia terbagi ke dalam dua *cluster*, yaitu *Cluster 0* dengan 29 provinsi yang memiliki konsumsi rendah-menengah dan *Cluster 1* dengan 5 provinsi berdaya konsumsi tinggi, khususnya wilayah Pulau Jawa. Nilai DBI sebesar 0,507 menunjukkan bahwa pengelompokan yang dihasilkan cukup optimal. Hasil ini sejalan dengan laporan PLN dan diharapkan dapat mendukung perencanaan infrastruktur serta kebijakan distribusi energi secara lebih tepat sasaran.

Kata Kunci: Data Mining; K-Means Clustering; Konsumsi Listrik

1. Pendahuluan

Penggunaan listrik menjadi bagian penting dalam kehidupan masyarakat Indonesia. Aktivitas sehari-hari seperti penerangan, memasak, dan penggunaan alat elektronik sangat

bergantung pada listrik. Perusahaan Listrik Negara (PLN) adalah perusahaan milik negara yang bergerak di bidang energi. PLN menyediakan listrik bagi seluruh warga, industri, dan sektor lainnya. Sejak tahun 1945, PLN telah berkembang pesat dalam menyediakan layanan kelistrikan bagi masyarakat Indonesia dan turut memainkan peran signifikan dalam mendukung pertumbuhan ekonomi nasional. Pada tahun 2023, Tingkat elektrifikasi di Indonesia pada tahun 2023 telah mencapai 99,74%, menunjukkan hampir seluruh wilayah telah teraliri listrik.[1].

Menurut laporan yang dirilis oleh Kementerian ESDM, konsumsi listrik per kapita di Indonesia mengalami kenaikan dari 1.173 kWh/kapita pada tahun 2022 menjadi 1.441 kWh/kapita pada tahun 2024. Selain itu, penjualan listrik PLN meningkat 5,32% dari tahun 2022 ke 2023, dengan segmen bisnis dan industri sebagai kontributor terbesar [2]. Penjualan listrik PT PLN (Persero) meningkat 5,32 persen dari 270,82 TWh pada tahun 2022 menjadi 285,23 TWh pada tahun 2023. Dengan pertumbuhan 12,53%, segmen bisnis menjadi penopang utama. Di sisi lain, segmen industri menyumbang 30,72% dari total penjualan energi. Peningkatan ini menunjukkan tren konsumsi listrik yang terus meningkat di Indonesia [3]. Meskipun tren konsumsi listrik terus meningkat, distribusi penggunaannya tidak merata di setiap provinsi. Wilayah dengan konsumsi listrik terbesar di Indonesia pada tahun 2023 adalah Jawa-Madura-Bali, dengan jumlah populasi mencapai 156 juta jiwa dan konsumsi listrik sebesar 175.035 GWh [4]. Mengingat banyaknya jumlah provinsi di Indonesia, rata-rata penjualan energi di setiap provinsi tentu berbeda-beda. Ketimpangan ini menunjukkan adanya perbedaan besar dalam penggunaan energi antar daerah. Masalah utamanya adalah belum adanya pengelompokan atau klasifikasi wilayah berdasarkan tingkat penggunaan listrik secara sistematis dan jelas. Sampai saat ini, PLN hanya memberikan data mentah berupa angka penjualan energi untuk tiap jenis pelanggan, tanpa proses analisis lebih lanjut yang bisa membantu menentukan wilayah yang paling membutuhkan dalam distribusi energi. Oleh karena itu, dibutuhkan pendekatan yang menganalisis pola penggunaan listrik di tiap wilayah agar bisa mengelompokkan daerah yang memiliki karakteristik yang serupa.

Untuk mengatasi masalah tersebut, diperlukan metode analisis data yang mampu mengelompokkan wilayah-wilayah dengan karakteristik konsumsi listrik yang serupa. Salah satu metode yang dapat digunakan adalah *K-Means Clustering*, metode dalam mengolah data yang termasuk dalam metode *unsupervised learning*. Algoritma ini digunakan untuk membagi data ke dalam beberapa *cluster* berdasarkan kesamaan sifat atau ciri yang dimiliki, dan memiliki cara kerja yang cepat serta mampu memberikan hasil pengelompokan dalam waktu singkat [5]. Metode ini dianggap efisien karena mampu mengolah data dalam skala besar serta menghasilkan kelompok yang sederhana namun tetap memberikan informasi yang bermakna. Pendekatan ini telah diterapkan dalam penelitian oleh Cici Astria *et al.* Dalam studi tersebut, tiga provinsi termasuk dalam *cluster* dengan tingkat distribusi tinggi, sedangkan 31 provinsi lainnya masuk ke dalam kategori distribusi rendah [6]. Dengan pengelompokan ini, pengambil kebijakan dapat lebih mudah menentukan daerah prioritas untuk pembangunan atau peningkatan infrastruktur kelistrikan.

Penelitian ini dilakukan untuk membagi provinsi di Indonesia menjadi beberapa kelompok berdasarkan pola penggunaan listrik dari berbagai kategori pelanggan dengan menerapkan metode *K-Means Clustering*. Hasil pengelompokan ini diharapkan bisa membantu PLN dan pemerintah dalam mengetahui wilayah mana saja yang tingkat konsumsi listriknya tinggi, atau rendah. Dengan begitu, perencanaan pembangunan infrastruktur listrik, pembagian daya listrik, dan kebijakan lainnya bisa dibuat dengan lebih tepat dan sesuai kebutuhan di setiap wilayah.

2. Tinjauan Pustaka

Pada penelitian sebelumnya membahas mengenai Data Mining yang diaplikasikan pada Kasus Tenaga Listrik yang Dibangkitkan menunjukkan bahwa algoritma *K-Means* efektif dalam mengelompokkan provinsi di Indonesia berdasarkan kapasitas pembangkitan listrik periode 2011–2017. Hasilnya, dua provinsi masuk kategori kapasitas tinggi, empat provinsi sedang, dan 27 provinsi rendah [7].

Penelitian kedua membuktikan bahwa metode *K-Means Clustering* berhasil diterapkan dalam menganalisa tingkat pemakaian listrik. Hasil penelitian menunjukkan perubahan dalam konsumsi, yang membantu mengidentifikasi area yang memerlukan intervensi, seperti kampanye konservasi energi [8].

Pada penelitian ketiga, dijelaskan bahwa pelanggan pascabayar PLN di Helvetia dikelompokkan berdasarkan biaya energi dan kode pembaca meteran menggunakan metode *K-means*. Hasil pengelompokan dibagi ke dalam tiga kategori yaitu tinggi, sedang, dan rendah. Hal ini dapat membantu PLN meningkatkan kualitas layanan dan mengembangkan strategi pemasaran, infrastruktur, dan program penghematan energi yang lebih terarah [9].

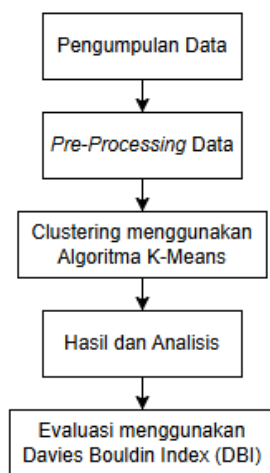
Penelitian lainnya membahas penerapan algoritma *K-Means* untuk mengklasifikasikan provinsi-provinsi di Indonesia berdasarkan jumlah listrik yang didistribusikan. Latar belakang penelitian ini berkaitan dengan meningkatnya kebutuhan energi dan tantangan dalam pemerataan distribusi. Data yang digunakan mencakup periode 2011–2017 dan mencakup 34 provinsi. Hasil analisis menghasilkan dua *cluster*, yaitu distribusi tinggi (C1) dan rendah (C2) [6].

Penelitian lain juga telah dilakukan dengan menggunakan algoritma Mini Batch *K-Means* Clustering untuk merancang sistem pemantauan penggunaan listrik pada gedung. Sistem berbasis unsupervised learning ini menampilkan informasi melalui website. Hasil uji menunjukkan nilai silhouette score pada Gedung N sebesar 0,62 (harian), 0,57 (bulanan), dan 0,73 (tahunan), menandakan struktur *cluster* yang baik [10].

Dari berbagai studi terdahulu, terlihat bahwa algoritma *K-Means* sangat efektif dalam mengelompokkan data berdasarkan kecocokan sifat-sifatnya. Proses pengelompokan ini memberikan informasi yang berguna untuk membantu membuat keputusan kebijakan yang lebih tepat dan akurat. Pada penelitian ini menawarkan pendekatan baru yaitu membagi provinsi di Indonesia berdasarkan tingkat pemakaian listrik berdasarkan jenis pelanggan. Penelitian ini tidak hanya mengandalkan total konsumsi, tetapi juga mempertimbangkan struktur konsumsi dari berbagai sektor pada tiap wilayah. Hal ini memberikan gambaran yang lebih lengkap mengenai pola konsumsi listrik secara menyeluruh. Selain itu, penelitian ini menggunakan data terbaru yang lebih relevan secara kebijakan (hingga tahun 2024) dan mencakup seluruh wilayah administratif (34 provinsi). Pendekatan ini memberikan kontribusi dalam membantu PLN dan pemangku kebijakan untuk memahami kebutuhan energi tiap wilayah secara lebih spesifik dan akurat, sehingga mendukung strategi perencanaan infrastruktur dan distribusi daya yang lebih efisien.

3. Metodologi

Penelitian ini dilakukan dengan menerapkan tahapan metode seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

1) Pengumpulan Data

Data yang digunakan dalam penelitian ini meliputi data Energi Terjual per Kelompok Pelanggan berdasarkan kelompok pelanggan dan wilayah di Indonesia, yang diperoleh dari sumber data resmi PT Perusahaan Listrik Negara (PLN). Periode data yang digunakan adalah tahun 2019 hingga tahun 2024. Parameter yang digunakan adalah provinsi, dan jenis pelanggan (rumah tangga, bisnis, industri, gedung pemerintah, dan penerangan jalan umum).

Dengan cakupan wilayah yang mencakup seluruh provinsi di Indonesia, data ini memberikan gambaran menyeluruh mengenai distribusi dan pola konsumsi listrik di berbagai provinsi.

2) *Pre-Processing Data*

Sebelum memulai proses *clustering*, data diolah terlebih dahulu agar siap digunakan. Tahapan ini meliputi seleksi atribut yang relevan, penghapusan wilayah gabungan yang berpotensi menyebabkan duplikasi, serta penyesuaian format angka, seperti menghapus pemisah ribuan agar dapat dikenali dan diproses dengan benar oleh RapidMiner. Serta, dilakukan akumulasi data setiap provinsi dari tahun 2019 hingga 2024 dengan menjumlahkan total energi terjual per provinsi untuk setiap tahun.

3) *Metode K-Means Clustering*

Setelah melewati proses *pre-processing* data dan data sudah siap digunakan, maka dilakukan *clustering*. Pada tahap ini, data diolah menggunakan RapidMiner. *RapidMiner* adalah *software* berbasis Java yang dirancang agar dapat dijalankan di berbagai sistem operasi, sehingga memberikan fleksibilitas bagi pengguna dalam melakukan analisis data di berbagai platform. Aplikasi ini menjadi salah satu solusi dalam penerapan analisis *data mining*, karena menyediakan teknik analisis deskriptif dan prediktif yang membantu pengguna dalam mengambil keputusan secara optimal [11]. Proses pengolahan dimulai dengan pemilihan atribut yang relevan, dilanjut dengan penerapan algoritma *K-Means Clustering*. *Clustering* adalah metode dalam data mining yang digunakan untuk membagi data menjadi beberapa kelompok berdasarkan kesamaan tertentu. *K-means* merupakan salah satu algoritma dalam metode *clustering* [12]. Proses pengelompokan data menggunakan algoritma *K-Means* melibatkan langkah-langkah berikut:

1. Inisialisasi *Centroid*

Sebagai langkah awal, jumlah *cluster* (c) ditentukan, kemudian beberapa titik *centroid* dipilih secara acak untuk memulai proses klasifikasi data.

2. *Clustering*

Kelompokkan setiap data ke dalam *cluster* yang memiliki pusat (*centroid*) terdekat. Jarak ini umumnya dihitung menggunakan jarak Euclidean, yaitu:

$$D(i,j) = \sqrt{[(X1i - X1j)^2 + (X2i - X2j)^2 + \dots + (Xki - Xkj)^2]} \quad (1)$$

Dimana:

- $D(i,j)$ = Jarak antara data ke- i dengan pusat *cluster* ke- j .
- Xki = Data ke- i pada atribut data ke- k
- Xkj = Titik pusat ke- j pada atribut ke- k

3. Pembaruan *Centroid* Kelompok

Untuk memperbarui posisi *centroid*, dihitunglah rata-rata dari semua data dalam setiap kelompok berdasarkan hasil klasifikasi yang sudah ada

$$uj(t+1) = (1/Nsj) * \sum_{j \in sj} xj \quad (2)$$

Dimana :

- $Uj(t+1)$ = *centroid* baru pada iterasi
- Nsj = jumlah data dalam *cluster* sj

4. Pengulangan

Proses *clustering* dan pembaruan *centroid* dilakukan berulang kali sampai hasil pengelompokan stabil atau jumlah iterasi maksimum telah terpenuhi.

5. Evaluasi Konvergensi

Lakukan evaluasi konvergensi dengan memantau perubahan posisi *centroid* atau memastikan atribusi data dalam *cluster* sudah stabil dalam beberapa iterasi.

6. Terminasi

Algoritma akan berhenti ketika kriteria penghentian tercapai, baik karena iterasi maksimum telah dicapai atau tidak ada perubahan signifikan dalam pengelompokan data [13].

4) Hasil dan Analisis

Setelah menerapkan algoritma *K-Means* pada data konsumsi listrik, diperoleh hasil pengelompokan menjadi 2 *cluster* utama. Kedua *cluster* ini merepresentasikan pola konsumsi yang berbeda. Visualisasi hasil *clustering* menggunakan *line chart* menggambarkan pembagian wilayah/provinsi ke dalam 2 *cluster* utama berdasarkan pola konsumsi energi listrik pada masing-masing sektor. Selain itu, visualisasi ini juga menampilkan perbandingan konsumsi energi berdasarkan jenis pelanggan. Analisis terhadap karakteristik tiap *cluster* mengungkapkan perbedaan signifikan dalam pola pemakaian listrik yang berkaitan dengan jenis pelanggan dan lokasi geografis.

5) Evaluasi Davies Bouldin Index (DBI)

Penilaian terhadap hasil *clustering* dilakukan dengan pendekatan *Davies-Bouldin Index* (DBI), yaitu suatu teknik evaluasi yang mengukur seberapa baik data terbagi dalam masing-masing *cluster*. Pendekatan pengujian DBI mencakup dua aspek, yaitu separasi dan kohesi. Kohesi merepresentasikan seberapa mirip data dalam suatu *cluster* terhadap pusat *clusternya*, sedangkan separasi menggambarkan seberapa jauh jarak antar pusat *cluster*. Pengelompokan yang ideal dicirikan oleh kohesi rendah dan pemisahan tinggi. Semakin dekat nilai Indeks Davis-Bouldin (DBI) ke 0, semakin akurat hasil pengelompokan. Semakin kecil nilai DBI, maka semakin optimal hasil pengelompokan yang diperoleh. [14].

4. Hasil dan Pembahasan

4.1 Pre-Processing Data

1. Data Selection:

Tahapan ini melibatkan pemilihan variabel yang dianggap relevan untuk dianalisis. Adapun variabel yang digunakan mencakup tahun (2019–2024), nama provinsi, serta kategori pelanggan.

Tabel 1. Data Energi Terjual berdasarkan Kelompok Pelanggan (Gwh) Tahun 2019 - 2024

Tahun	Satuan PLN/ Provinsi	Rumah Tangga	Industri	Bisnis	Sosial	Gdg Kantor Pemerintahan	Penerangan Jalan Umum
2019	UIW Aceh	1733,04	159,59	466,06	199,25	104,06	119,5
2019	UIW North Sumatera	5362,88	2910,53	1695,35	422,76	141,08	411,26
2019	UIW West Sumatera	1650,11	925,12	535,31	165,07	76,81	92,66
2019	UIW Riau and Riau Archipelago	3091,53	486,04	1395,39	248,15	142,13	169,81
2019	Riau	2635,8	447,31	1098,76	213,52	103,04	148,36
2019	Riau Archipelago	455,73	38,73	296,62	34,63	39,09	21,45
2019	UIW South Sumatera, Jambi, and Bengkulu	4829,39	1195,37	1445,64	309,42	183,18	178,54
2019	South Sumatera	2923,6	969,99	925,91	195,82	106,07	132,67
2019	Jambi	1236,1	151,43	386,29	75,82	46,76	35,6
2019	Bengkulu	669,68	73,95	133,43	37,77	30,36	10,28
2019	UIW Bangka Belitung	689,04	185,66	204,76	43,87	34,42	9,18
2019	UID Lampung	2745,52	974,28	587,24	184,92	71,23	122,9
2019	UIW West Kalimantan	1588,66	183,33	567,38	107,99	78,52	46,81
2019	UIW South and Central Kalimantan	2508,13	448,58	825,75	175,77	138,61	81,07
2019	South Kalimantan	1659,4	358,77	543,08	120,92	77,2	59,79
2019	Kalimantan Central	848,73	89,82	282,67	54,85	61,41	21,29

Tahun	Satuan PLN/ Provinsi	Rumah Tangga	Industri	Bisnis	Sosial	Gdg Kantor Pemerintahan	Penerangan Jalan Umum
	Kalimantan						
2019	UIW East and North	2213,45	312,89	1020,3	182,28	170,19	53,77
2019	UIW North, Central Sulawesi, and Gorontalo	1967,25	387,26	692,59	162,22	136,54	125,38
...	...	...	...	...	...	...	...
2024	UID Jakarta Raya	16413,87	3915,90	13807,54	1742,43	1452,01	198,1

2. Data Cleaning

Tahapan ini mencakup proses pembersihan terhadap data yang akan dianalisis, meliputi penghapusan pemisah ribuan (seperti angka titik atau koma) serta penghapusan nilai gabungan pada data provinsi agar dapat diproses dengan benar oleh *Rapid Miner*.

Tabel 2. Menghapus Pemisah Ribuan

Total (Sebelum)	Total (Sesudah)
8,332.54	8322.54
5,308.45	5308.45
2,010.60	2010.60
1,013.49	1013.49
1,217.51	1217.51
4,959.02	4959.02
2,715.67	2715.67

Tabel 3. Menghapus Nilai Gabungan

Provinsi (Sebelum)	Provinsi (Sesudah)
UID Sumatera Selatan, Jambi, Bengkulu	-
Sumatera Selatan	Sumatera Selatan
Jambi	Jambi
Bengkulu	Bengkulu
UID Riau dan Kepulauan Riau	-
Riau	Riau
Kepulauan Riau	Kepulauan Riau

3. Data Transformasi

Pada tahap transformasi data, dilakukan proses penggabungan data dari seluruh tahun yang tersedia menjadi satu kesatuan. Proses ini bertujuan untuk mengakumulasi total nilai pada setiap kategori atau atribut yang dianalisis, sehingga menghasilkan representasi data yang lebih ringkas dan relevan. Dengan cara ini, pola konsumsi atau tren berdasarkan wilayah dapat dianalisis secara keseluruhan tanpa terfragmentasi oleh periode waktu tertentu.

Tabel 4. Hasil Akumulasi Data

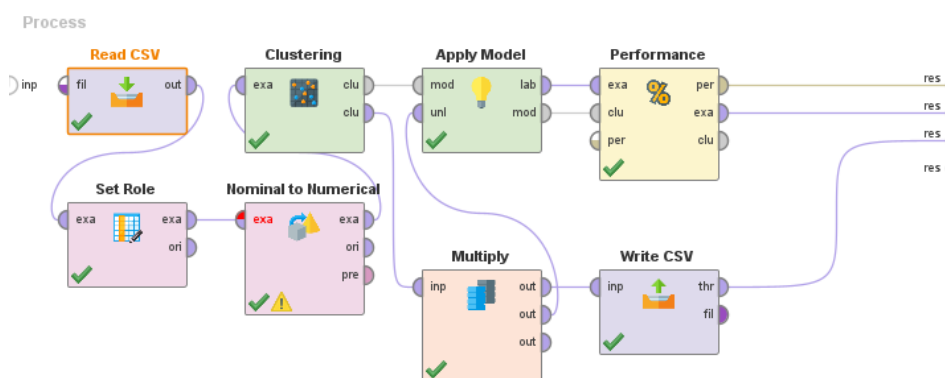
Satuan PLN/Provinsi	Rumah Tangga	Industri	Bisnis	Sosial	Gedung Kantor Pemerintahan	Penerangan Jalan Umum
Aceh	11553.63	1434.09	3310.34	1356.92	660.11	774.08
Sumatera Utara	35834.71	18512.87	10739.48	2955.14	909.18	2494.58
Sumatera Barat	10681.05	5222.04	3772.14	1136.22	491.15	579.89

Satuan PLN/Provinsi	Rumah Tangga	Industri	Bisnis	Sosial	Gedung Kantor Pemerintahan	Penerangan Jalan Umum
Riau	17899.8	12231.78	7317.38	1532.19	681.72	737.14
Kepulauan Riau	3023.01	239.65	1801.86	242.9	248.39	137.37
Sumatera Selatan	19697.58	6178.47	5908.07	1363.08	687.77	832.2
Jambi	8447.02	1051.78	2546.5	575.98	302.81	238.76
Bengkulu	4566.26	478.29	939.2	269.13	188.1	56.32
Bangka Belitung	4579.27	1919.57	1340.43	320.42	217.63	57.15
Lampung	18467.66	6599.8	4329.55	1353.31	442.42	750.23
Kalimantan Barat	11336.01	1399.05	3773.25	800.86	501.05	284.2
Kalimantan Selatan	11134.31	3290.43	3536.19	890.76	507.68	381.47
Kalimantan Tengah	6038.76	959.25	1973.27	443.09	389.56	137.26
Kalimantan Timur	15224.44	3487.42	6677.22	1271.67	1102.01	322.35
Sulawesi Utara	5781.49	2302.58	2601.92	540.02	302.24	256.03
Sulawesi Tengah	3807.92	461.06	892.9	296.07	283.25	103.46
Gorontalo	4448.7	403.42	963.87	335.72	294.8	251.33
Sulawesi Selatan	20025.34	12505.95	7391.47	1916.86	985.21	801.03
Sulawesi Tenggara	4723.05	547.21	1375.49	317.26	275.82	121.49
Sulawesi Barat	1850.12	231.4	397.76	138.81	135.19	45.42
Maluku	2102.17	66.2	845.64	202.37	286.5	45.65
Maluku Utara	2751.26	857.02	689.86	190.55	238.68	60.53
Papua	4432.17	65.04	1941.4	434.67	535.41	71.96
Papua Barat	2175.59	63.3	885.43	213.34	273.13	37.24
Bali	15507.44	1205.24	15459.39	1115.96	696.18	398.82
Nusa Tenggara Barat	9056.69	1186.29	2639.81	635.15	293.96	339.16
Nusa Tenggara Timur	4540.08	251.51	1612.5	451.05	306.21	102.22
Batam	5619.61	5331.92	5430.86	467.22	251.32	84.68
Jawa Timur	93496.19	99227.73	32705.89	9093.53	2687.98	3611.13
Jawa Tengah	78145.25	50192.95	22197.53	7330.79	1824.95	2942.16
DI Yogyakarta	10901.85	1602.01	4881.75	1725.31	446.93	376.44
Jawa Barat	125686.22	145579.32	44369.5	8570.98	2964.28	2311.28
Banten	36413.32	90786.14	20316.25	2350.79	983.03	508.72
Jakarta Raya	90179.6	24418.89	75012.63	8796.31	8391	1286.08

4.2 K-Means Clustering

Clustering dilakukan dengan menerapkan algoritma *K-Means* untuk mengelompokkan data berdasarkan kesamaan karakteristiknya. Tujuan dari proses ini adalah untuk mengelompokkan wilayah atau provinsi berdasarkan kesamaan pola penggunaan energi listrik di setiap sektor. *Clustering* dibagi menjadi 2 *cluster*. Proses *clustering* ini dilakukan setelah data melalui tahapan *cleaning* dan transformasi, sehingga data yang digunakan telah siap dianalisis secara optimal.

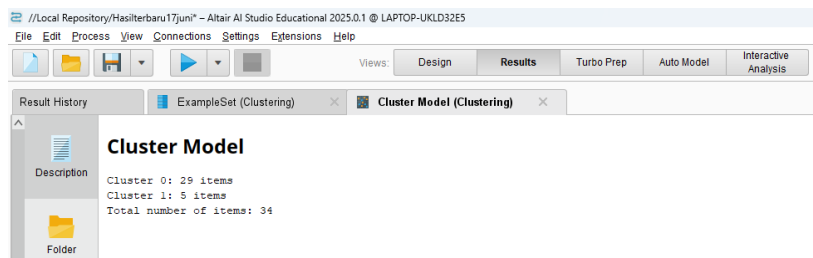
1. Proses Modelling



Gambar 2. Proses Modelling

2. Hasil Clustering

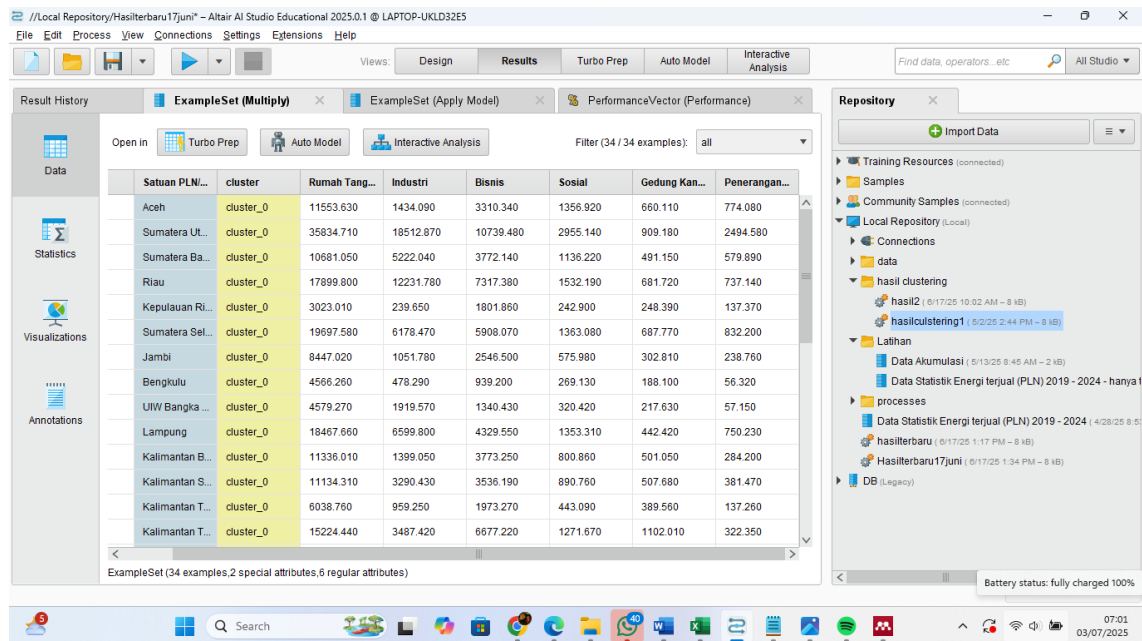
Proses pemodelan dengan algoritma *K-Means Clustering* menghasilkan dua kelompok data, yaitu *cluster 0* dan *cluster 1*. Masing-masing *cluster* merepresentasikan kelompok wilayah dengan karakteristik konsumsi energi listrik yang serupa berdasarkan sektor penggunaannya.



Gambar 3 Jumlah Provinsi Setiap Cluster

Attribute	cluster_0	cluster_1
Rumah Tangga	9524.379	84784.116
Industri	3106.367	82041.006
Bisnis	3654.308	38920.360
Sosial	810.070	7228.480
Gedung Kantor Pemerintahan	446.014	3370.248
Penerangan Jalan Umum	375.119	2131.874

Gambar 4 Data Centroid



Gambar 5. Hasil Clustering

Tabel 5 Hasil Cluster 0

Cluster	Provinsi	Rumah Tangga	Industri	Bisnis	Sosial	Gedung Kantor	Penerangan Jalan Umum
0	Aceh	11553.63	1434.09	3310.34	1356.92	660.11	774.08
0	Sumatera Utara	35834.71	18512.87	10739.48	2955.14	909.18	2494.58
0	Sumatera Barat	10681.05	5222.04	3772.14	1136.22	491.15	579.89
0	Riau	17899.8	12231.78	7317.38	1532.19	681.72	737.14
0	Kepulauan Riau	3023.01	239.65	1801.86	242.9	248.39	137.37
0	Sumatera Selatan	19697.58	6178.47	5908.07	1363.08	687.77	832.2
0	Jambi	8447.02	1051.78	2546.5	575.98	302.81	238.76
0	Bengkulu	4566.26	478.29	939.2	269.13	188.1	56.32
0	Bangka Belitung	4579.27	1919.57	1340.43	320.42	217.63	57.15
0	Lampung	18467.66	6599.8	4329.55	1353.31	442.42	750.23
0	Kalimantan Barat	11336.01	1399.05	3773.25	800.86	501.05	284.2
0	Kalimantan Selatan	11134.31	3290.43	3536.19	890.76	507.68	381.47
0	Kalimantan Tengah	6038.76	959.25	1973.27	443.09	389.56	137.26
0	Kalimantan Timur	15224.44	3487.42	6677.22	1271.67	1102.01	322.35
0	Sulawesi Utara	5781.49	2302.58	2601.92	540.02	302.24	256.03
0	Sulawesi Tengah	3807.92	461.06	892.9	296.07	283.25	103.46
0	Gorontalo	4448.7	403.42	963.87	335.72	294.8	251.33
0	Sulawesi Selatan	20025.34	12505.95	7391.47	1916.86	985.21	801.03
0	Sulawesi Tenggara	4723.05	547.21	1375.49	317.26	275.82	121.49
0	Sulawesi Barat	1850.12	231.4	397.76	138.81	135.19	45.42

Cluster	Provinsi	Rumah Tangga	Industri	Bisnis	Sosial	Gedung Kantor	Penerangan Jalan Umum
0	Maluku	2102.17	66.2	845.64	202.37	286.5	45.65
0	Maluku Utara	2751.26	857.02	689.86	190.55	238.68	60.53
0	Papua	4432.17	65.04	1941.4	434.67	535.41	71.96
0	Papua Barat	2175.59	63.3	885.43	213.34	273.13	37.24
0	Bali	15507.44	1205.24	15459.39	1115.96	696.18	398.82
0	Nusa Tenggara Barat	9056.69	1186.29	2639.81	635.15	293.96	339.16
0	Nusa Tenggara Timur	4540.08	251.51	1612.5	451.05	306.21	102.22
0	Batam	5619.61	5331.92	5430.86	467.22	251.32	84.68
0	DI Yogyakarta	10901.85	1602.01	4881.75	1725.31	446.93	376.44

Tabel 6 Hasil *Cluster 1*

Cluster	Provinsi	Rumah Tangga	Industri	Bisnis	Sosial	Gedung Kantor	Penerangan Jalan Umum
1	Jawa Timur	93496.19	99227.73	32705.89	9093.53	2687.98	3611.13
1	Jawa Tengah	78145.25	50192.95	22197.53	7330.79	1824.95	2942.16
1	Jawa Barat	125686.22	145579.32	44369.5	8570.98	2964.28	2311.28
1	Banten	36413.32	90786.14	20316.25	2350.79	983.03	508.72
1	Jakarta Raya	90179.6	24418.89	75012.63	8796.31	8391	1286.08

3. Interpretasi Hasil *cluster*

Tabel 7 Interpretasi *Cluster 0*

Jenis Pelanggan	Tertinggi	Provinsi	Terendah	Provinsi
Rumah Tangga	35834.71	Sumatra Utara	1850.12	Sulawesi Barat
Industri	18512.87	Sumatra Utara	66.2	Maluku
Bisnis	15459.39	Bali	397.76	Sulawesi Barat
Sosial	2955.14	Sumatra Utara	138.81	Sulawesi Barat
Gedung Kantor Pemerintahan	1102.01	Kalimantan Timur	135.19	Sulawesi Barat
Penerangan Jalan Umum	2494.58	Sumatra Utara	37.24	Papua Barat

Berdasarkan hasil pengelompokan dengan algoritma *K-Means*, *Cluster 0* mencakup provinsi-provinsi dengan konsumsi listrik yang rendah di hampir semua jenis pelanggan. Sumatera Utara menjadi pengecualian karena mencatat nilai tertinggi pada sektor rumah tangga, industri, sosial, dan penerangan jalan umum. Sebaliknya, provinsi seperti Sulawesi Barat, Papua Barat, dan Maluku menunjukkan konsumsi sangat rendah, terutama pada sektor bisnis dan gedung kantor. Secara umum, *Cluster 0* merepresentasikan wilayah dengan tingkat pemakaian listrik yang masih rendah dan berpotensi untuk ditingkatkan.

Tabel 8 Interpretasi *Cluster 1*

Jenis Pelanggan	Tertinggi	Provinsi	Terendah	Provinsi
Rumah Tangga	125686.22	Jawa Barat	36413.32	Banten
Industri	145579.32	Jawa Barat	24418.89	Jakarta Raya
Bisnis	75012.63	Jakarta Raya	44369.5	Jawa Barat
Sosial	9093.53	Jawa Timur	2350.79	Banten
Gedung Kantor Pemerintahan	8391	Jakarta Raya	983.03	Banten
Penerangan Jalan Umum	3611.13	Jawa Timur	508.72	Banten

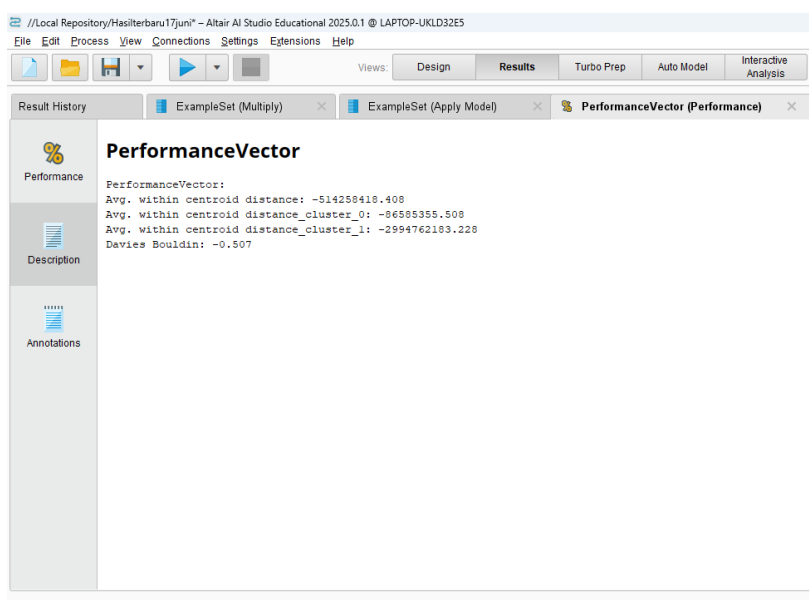
Berdasarkan hasil pengelompokan *Cluster 1*, wilayah-wilayah di dalamnya umumnya menunjukkan konsumsi listrik yang tinggi dan merata, terutama pada sektor rumah tangga, industri, dan gedung pemerintahan. Hal ini tercermin dari dominasi provinsi besar seperti Jawa Barat, Jakarta Raya, dan Jawa Timur yang menduduki peringkat tertinggi pada sebagian besar kategori pelanggan.

4. Evaluasi / Pengujian menggunakan *Davies Bouldin Index*

Hasil Evaluasi menunjukkan *cluster* terbaik yaitu 2 dengan nilai 0,508. Nilai ini menunjukkan bahwa hasil pengelompokan yang diperoleh melalui algoritma *K-Means* tergolong optimal, karena nilai DBI yang rendah mencerminkan adanya pemisahan yang jelas antar *cluster* dan kemiripan yang tinggi di dalam masing-masing *cluster*. Dengan demikian, model *clustering* yang dibentuk dapat dikatakan cukup optimal dalam mengelompokkan data berdasarkan pola konsumsi energi listrik.

Tabel 9 Hasil Nilai *Davies Bouldin Index* (DBI)

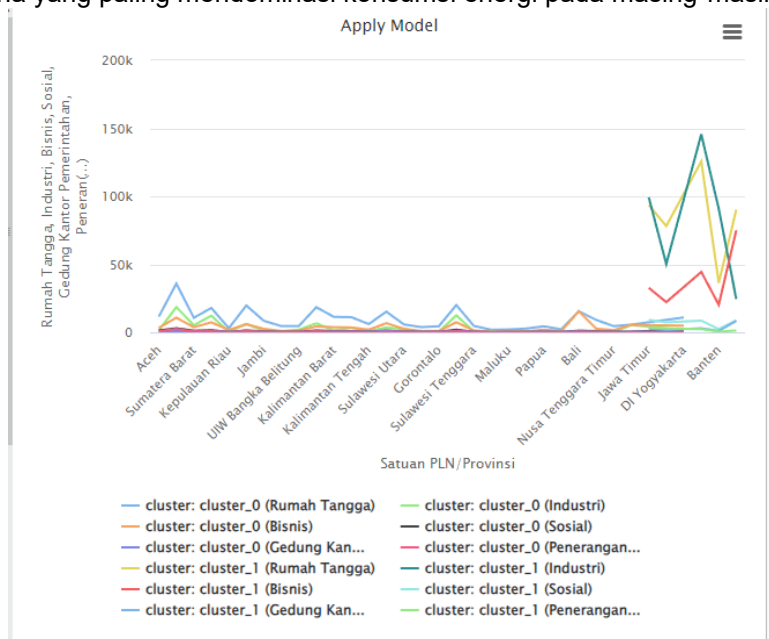
Penentuan Jumlah <i>Cluster</i> Terbaik	Hasil DBI
2 <i>Cluster</i>	0,507
3 <i>Cluster</i>	0,781
4 <i>Cluster</i>	0,584

Gambar 6. *Performance Vector*

5. Visualisasi *Clustering*

Hasil visualisasi di bawah ini menunjukkan pengelompokan data yang diperoleh melalui penerapan algoritma *K-Means*. Visualisasi ini menggambarkan pembagian wilayah/provinsi ke dalam 2 *cluster* utama berdasarkan pola konsumsi energi listrik pada masing-masing sektor. Selain itu, visualisasi ini juga menampilkan perbandingan konsumsi energi berdasarkan jenis

pelanggan, sehingga memudahkan dalam menginterpretasi hasil pengelompokan secara visual dan sektor mana yang paling mendominasi konsumsi energi pada masing-masing *cluster*.



Gambar 7. Visualisasi *Clustering*

4.3 Pengujian Akurasi Algoritma

Tabel 10 Perbandingan Hasil *Cluster* Berdasarkan Provinsi

Provinsi	Hasil <i>Cluster</i>	Data PLN	Sesuai atau Tidak
Pulau Jawa (Jakarta Raya, Jawa Barat, Jawa Timur, Jawa Tengah, Banten)	Tinggi	Tinggi [15]	Sesuai
Luar Pulau Jawa	Lebih rendah dibanding pulau Jawa	Lebih rendah dibanding pulau Jawa [15]	Sesuai

Tabel menunjukkan bahwa provinsi di Pulau Jawa (seperti Jakarta, Jawa Barat, dan lainnya) termasuk dalam *cluster* dengan konsumsi energi tinggi, sesuai dengan data PLN. Sementara itu, provinsi di luar Pulau Jawa tergolong dalam *cluster* dengan konsumsi lebih rendah, juga sesuai dengan data PLN. Hal ini menunjukkan bahwa hasil *clustering* sejalan dengan kondisi riil distribusi energi pada data PLN.

4.4 Pembahasan

Penelitian ini memperkuat hasil-hasil dari penelitian sebelumnya yang juga menggunakan algoritma *K-Means* dalam mengelompokkan data konsumsi energi listrik. Beberapa studi terdahulu telah menunjukkan bahwa *K-Means* dapat digunakan secara efektif untuk mengidentifikasi pola konsumsi berdasarkan wilayah atau jenis pelanggan. Pada penelitian yang membahas tentang pendistribusian listrik, berhasil dikelompokkan dengan algoritma *K-Means Clustering*, hasilnya *clustering* membagi wilayah ke dalam dua *cluster*, yaitu distribusi tinggi dan rendah [6]. Penelitian ini memiliki kemiripan dalam penggunaan variabel wilayah, namun cakupannya terbatas pada satu aspek yaitu volume distribusi. Berbeda dari penelitian tersebut, penelitian ini memberikan kontribusi, yaitu:

1. Menggunakan lebih banyak atribut, seperti jenis-jenis pelanggan (Rumah Tangga, Industri, Bisnis, Sosial, Gedung Kantor, dan Penerangan Jalan Umum)

2. Menggunakan data yang lebih terbaru, yaitu data konsumsi energi listrik dari tahun 2019 hingga tahun 2024, sementara penelitian sebelumnya hanya mencakup data hingga tahun 2017.
 3. Melakukan evaluasi dengan *Davies-Bouldin Index* (DBI) untuk mengukur performa model
- Penelitian ini tidak hanya memperluas cakupan wilayah, tetapi juga menyajikan analisis berdasarkan jenis-jenis pelanggan, sehingga memberikan informasi yang lebih spesifik mengenai sektor mana yang memiliki konsumsi energi tertinggi di masing-masing *cluster*.

5. Simpulan

Algoritma *K-Means Clustering* berhasil mengelompokkan 34 provinsi di Indonesia menjadi 2 *cluster*, berdasarkan data konsumsi energi listrik dari berbagai jenis pelanggan, seperti rumah tangga, industri, bisnis, sosial, gedung kantor pemerintahan, dan penerangan jalan umum. Hasil pengelompokan menunjukkan bahwa *Cluster 0* terdiri dari 29 provinsi dengan tingkat konsumsi listrik yang relatif rendah, sementara *Cluster 1* mencakup 5 provinsi dengan konsumsi tinggi. Evaluasi menggunakan *Davies-Bouldin Index* (DBI) menghasilkan nilai 0,507 yang menunjukkan bahwa hasil pengelompokan cukup optimal dengan struktur *cluster* yang baik. Temuan ini juga selaras dengan laporan konsumsi listrik PLN, di mana wilayah Jawa mendominasi konsumsi nasional. Selain itu, penelitian ini memberikan kontribusi baru dengan memanfaatkan data yang lebih terkini (tahun 2019–2024), mempertimbangkan variasi jenis pelanggan secara lebih detail, serta mengukur performa model *clustering*. Diharapkan penelitian ini dapat memberikan hasil yang dapat dijadikan acuan dalam perencanaan penyaluran energi dan pengambilan keputusan yang lebih tepat bagi PLN dan pemerintah.

Daftar Referensi

- [1] PLN, "78 Tahun PLN Hadir Listriki Indonesia, Rasio Elektrifikasi Nasional Capai 99,74%," PLN. Accessed: Oct. 20, 2024. [Online]. Available: <https://web.pln.co.id/media/siaran-pers/2023/10/78-tahun-pln-hadir-listriki-indonesia-rasio-elektrifikasi-nasional-capai-9974>
- [2] V. N. Setiawan, "Konsumsi Listrik RI di 2024 Naik Jadi 1.411 kWh/Kapita," CNBC INDONESIA. [Online]. Available: <https://www.cnbcindonesia.com/news/20250203151548-4-607561/konsumsi-listrik-ri-di-2024-naik-jadi-1411-kwh-kapita>
- [3] G. A. Trianto, "Penjualan Meningkat, Ini Upaya PLN Listriki Sektor Bisnis dan Industri Sepanjang Tahun 2023," PLN. Accessed: Oct. 21, 2024. [Online]. Available: <https://web.pln.co.id/media/siaran-pers/2024/01/penjualan-meningkat-ini-upaya-pln-listriki-sektor-bisnis-dan-industri-sepanjang-tahun-2023>
- [4] A. H. Wibowo, "Perbandingan Konsumsi Listrik Pulau-Pulau Besar di Indonesia," GoodStats. Accessed: Oct. 20, 2024. [Online]. Available: <https://data.goodstats.id/statistic/perbandingan-konsumsi-listrik-pulau-pulau-besar-di-indonesia-imT74>
- [5] Y. P. Anggriani, A. Arif, and F. Febriansyah, "Implementasi Algoritma K-Means Clustering Dalam Menentukan Blok Tanaman Sawit Produktif Pada Pt Arta Prigel," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1820–1825, 2024, doi: 10.36040/jati.v8i2.9225.
- [6] C. Astria, A. P. Windarto, A. Wanto, and E. Irawan, "Metode K-Means pada Pengelompokan Wilayah Pendistribusian Listrik," *Semin. Nas. Sains Teknol. Inf.*, pp. 306–312, 2019, [Online]. Available: <http://prosiding.seminar-id.com/index.php/sensasi/issue/archivePage%7C306>
- [7] A. Wati, I. Indriani, T. S. S. Manihuruk, S. Sintya, I. Y. Manurung, and A. P. Windarto, "Implementasi Datamining Pada Kasus Tenaga Listrik Yang Dibangkitkan Berdasarkan Provinsi," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 719–727, 2019, doi: 10.30865/komik.v3i1.1683.
- [8] R. D. Herdiansyah, "*Penerapan Algoritma K-Means Clustering Pada Penentuan Tingkat Pemakaian Listrik (Studi Kasus: PLTBg Rantau Sakti)*," Skripsi, Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim, 2023.
- [9] I. F. Batubara and F. Z. Lubis, "Clustering Data Pelanggan Pln Helvetia Menggunakan Metode K-Means Cluster," *J. Multidisiplin Saintek*, vol. 2, no. 1, pp. 71–80, 2023, [Online]. Available: <https://ejournal.warunayama.org/kohesi>
- [10] A. A. Herlambang, M. A. Murti, and C. Setianingsih, "Pengelompokan Data Penggunaan Energi Listrik menggunakan Algoritma Mini Batch K-Means Clustering," *e-Proceeding Eng.*, vol. 9, no. 5, pp. 2144–2153, 2022.

-
- [11] A. Apriyadi, M. R. Lubis, and B. E. Damanik, "Penerapan Algoritma C5.0 Dalam Menentukan Tingkat Pemahaman Mahasiswa Terhadap Pembelajaran Daring," *Komputa J. Ilm. Komput. dan Inform.*, vol. 11, no. 1, pp. 11–20, 2022, doi: 10.34010/komputa.v11i1.7386.
- [12] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024, [Online]. Available: <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- [13] S. Suraya, M. Sholeh, and U. Lestari, "Evaluation of Data Clustering Accuracy using K-Means Algorithm," *Int. J. Multidiscip. Approach Res. Sci.*, vol. 2, no. 01, pp. 385–396, 2023, doi: 10.59653/ijmars.v2i01.504.
- [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," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.
- [15] IESR, "Menerangi Indonesia," *Pojok Energi*, 2017.