

Implementation of the K-Means Algorithm on Household Electric Load

Implementasi Algoritma K-Means Pada Beban Listrik Rumah Tangga

Alief Nur Aisyi Maulidhia¹, Indri Ika Widyastuti², Friska Intan Sukarno³

Rahmat Basya Shahrys Tsany⁴ Thomas Brian⁵

^{1,5}Prodi Teknik Kelistrikan Kapal, Politeknik Perkapalan Negeri Surabaya ²Prodi Teknik

Bangunan Kapal, Politeknik Perkapalan Negeri Surabaya, ³Prodi Manajemen

Bisnis, Politeknik Perkapalan Negeri Surabaya

⁴Prodi Teknik Perpipaan, Politeknik Perkapalan Negeri Surabaya

Politeknik Perkapalan Negeri Surabaya

E-mail: *¹aliefnur@ppns.ac.id, ²indriwidyastuti@ppns.ac.id, ³friskaintan@ppns.ac.id

⁴rahmatbasya@ppns.ac.id, ⁵thomasbrian@ppns.ac.id

Abstract – The increase in household electricity consumption demands a more efficient power management strategy. One approach that can be used to analyze electricity consumption patterns is by applying the K-Means Clustering algorithm. This algorithm enables the grouping of electricity consumption data based on similarities in power usage patterns, helping to understand household energy consumption behavior and providing recommendations for energy efficiency.

This study aims to implement the K-Means algorithm on household electricity consumption data collected at one-minute intervals over a specific period. The data used includes various electrical parameters such as active power, reactive power, voltage, and sub-metering values from several electrical appliances. The research stages include data collection and preprocessing, the application of the K-Means algorithm for clustering power consumption patterns, and the evaluation of clustering results. The results of the study indicate that the K-Means algorithm can successfully classify household electricity consumption patterns into several categories based on power usage levels. This classification helps users identify high-energy-consuming appliances and develop energy-saving strategies. Furthermore, the analysis results can be utilized by electricity service providers to optimize power distribution and design more efficient policies. The optimal research result shows the number of clusters = 3.

Keywords — K-Means Clustering, household electricity consumption, load patterns, energy efficiency.

Abstrak – Peningkatan konsumsi energi listrik rumah tangga menuntut adanya strategi pengelolaan daya yang lebih efisien. Salah satu pendekatan yang dapat digunakan untuk menganalisis pola konsumsi listrik adalah dengan menerapkan algoritma *K-Means Clustering*. Algoritma ini memungkinkan pengelompokan data konsumsi listrik berdasarkan kesamaan pola penggunaan daya, sehingga dapat membantu dalam memahami perilaku konsumsi energi rumah tangga serta memberikan rekomendasi efisiensi listrik. Penelitian ini bertujuan untuk mengimplementasikan algoritma *K-Means* pada data konsumsi listrik rumah tangga yang dikumpulkan dengan interval satu menit selama periode tertentu. Data yang digunakan mencakup berbagai parameter kelistrikan seperti daya aktif, daya reaktif, tegangan, dan nilai *sub-metering* dari beberapa perangkat listrik. Tahapan penelitian meliputi pengumpulan dan pra-proses data, penerapan algoritma *K-Means* untuk pengelompokan pola konsumsi daya, serta evaluasi hasil *clustering*. Hasil penelitian menunjukkan bahwa algoritma *K-Means* mampu mengelompokkan pola konsumsi listrik rumah tangga menjadi beberapa kategori berdasarkan tingkat penggunaan daya. Pengelompokan ini dapat membantu pengguna dalam mengidentifikasi perangkat dengan konsumsi energi tinggi serta merancang strategi penghematan energi. Selain itu, hasil analisis dapat digunakan oleh penyedia layanan listrik untuk mengoptimalkan distribusi daya dan merancang kebijakan yang lebih efisien. Hasil penelitian optimal menunjukkan jumlah *cluster*=3

Kata Kunci — *K-Means Clustering*, konsumsi listrik rumah tangga, pola beban listrik, efisiensi energi.

1. PENDAHULUAN

Konsumsi energi listrik rumah tangga terus meningkat seiring dengan bertambahnya jumlah perangkat elektronik yang digunakan dalam kehidupan sehari-hari. Pengelolaan konsumsi listrik yang tidak efisien dapat menyebabkan pemborosan energi serta peningkatan biaya listrik bagi pengguna. Oleh karena itu, diperlukan metode yang efektif untuk menganalisis dan mengelompokkan pola konsumsi daya listrik rumah tangga guna membantu optimalisasi penggunaan energi [1]. Pemborosan energi listrik memiliki dampak yang signifikan, baik secara lingkungan, ekonomi, maupun sosial. Secara lingkungan, pemborosan listrik meningkatkan beban pada pembangkit listrik, terutama yang berbahan bakar fosil, sehingga menghasilkan lebih banyak emisi gas rumah kaca seperti CO₂ yang memperparah perubahan iklim, serta polutan seperti sulfur dioksida yang menyebabkan hujan asam dan masalah kesehatan. Dari segi ekonomi, pemborosan energi menyebabkan biaya listrik naik karena permintaan yang tinggi, membebani infrastruktur jaringan listrik, dan meningkatkan risiko pemadaman. Secara sosial, pemborosan energi dapat memperlebar ketimpangan akses energi, di mana masyarakat kurang mampu kesulitan mendapatkan listrik karena sumber daya yang terbatas. Contoh sederhana seperti membiarkan lampu menyala di ruangan kosong atau menggunakan AC berlebihan tidak hanya meningkatkan tagihan listrik, tetapi juga berkontribusi pada kerusakan lingkungan jangka panjang.

Salah satu metode yang dapat digunakan untuk mengidentifikasi dan mengelompokkan pola konsumsi listrik adalah algoritma *K-Means Clustering*. Algoritma *K-Means* dipilih dalam judul penelitian "*Implementasi Algoritma K-Means Pada Beban Listrik Rumah Tangga*" karena kemampuannya dalam melakukan *clustering* (pengelompokan) secara efektif dan efisien. Algoritma ini mampu mengelompokkan data konsumsi listrik rumah tangga berdasarkan pola penggunaan daya, seperti tinggi, sedang, dan rendah, dengan cara yang sederhana dan mudah diinterpretasikan. *K-Means* juga dikenal skalabel untuk *dataset* berukuran sedang hingga besar, cocok untuk menganalisis data konsumsi listrik yang melibatkan banyak rumah tangga. Selain itu, algoritma ini telah banyak digunakan dalam penelitian sejenis, seperti analisis pola konsumsi energi, sehingga telah teruji kehandalannya. Dengan menentukan jumlah *cluster* (*K*) sesuai kebutuhan, *K-Means* memungkinkan peneliti untuk memahami kebiasaan penggunaan listrik dan merancang strategi penghematan energi secara lebih terarah. Metode Penelitian Algoritma *K-Means Clustering* adalah metode *unsupervised learning* yang digunakan untuk mengelompokkan data ke dalam *k* kluster berdasarkan kesamaan karakteristik. Algoritma ini bekerja dengan menentukan sejumlah kluster awal (*k*) secara acak [3], kemudian menghitung jarak setiap data ke pusat kluster menggunakan *Euclidean Distance*. Selanjutnya, data diklasifikasikan ke dalam kluster dengan jarak terdekat, dan pusat kluster diperbarui berdasarkan rata-rata posisi data dalam kluster tersebut. Proses ini diulang hingga kluster mencapai keadaan stabil atau konvergen, di mana tidak ada perubahan signifikan dalam pembagian kelompok. *K-Means* banyak digunakan dalam berbagai aplikasi seperti analisis pola konsumsi energi, segmentasi pelanggan, dan pengelompokan gambar karena kemampuannya dalam menangani data dalam jumlah besar secara efisien [4].

Pada penelitian ini menggunakan *dataset Household Electric Power Consumption* yang diperoleh dari situs *Kaggle.com*, Pemilihan *dataset* dari *Kaggle* didasarkan pada beberapa alasan utama, yaitu ketersediaan data yang beragam, kredibilitas sumber, dan kemudahan akses. *Kaggle* menyediakan ribuan *dataset* yang telah diverifikasi dan digunakan oleh komunitas data scientist global, sehingga kualitas datanya terjamin. Selain itu, *dataset* di *Kaggle* sering dilengkapi dengan dokumentasi yang jelas, termasuk deskripsi variabel dan tujuan penggunaannya, yang memudahkan peneliti dalam memahami dan memproses data. *Platform* ini juga menyediakan fitur untuk berdiskusi dengan pengguna lain, sehingga peneliti dapat mendapatkan wawasan tambahan atau solusi jika menemui kendala dalam penggunaan *dataset* yang merupakan situs yang menyediakan beragam *dataset* [10].

2. METODE PENELITIAN

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari pengukuran konsumsi listrik rumah tangga dengan interval satu menit dalam periode tertentu. Data yang disajikan pada Tabel 1 ini mencakup beberapa parameter kelistrikan, antara lain:

Tabel 1. Deskripsi Fitur

No	Atribut	Deskripsi
1	<i>Date</i>	Tanggal dalam format dd/mm/yyyy
2	<i>Time</i>	Waktu dalam format hh:mm:ss
3	<i>Global Active Power</i>	Daya aktif rata-rata per menit secara global pada rumah tangga (dalam kilowatt)
4	<i>Global Reactive Power</i>	Daya reaktif rata-rata per menit secara global pada rumah tangga (dalam kilowatt)
5	<i>Voltage</i>	Tegangan rata-rata per menit (dalam volt)
6	<i>Global Intensity</i>	Intensitas arus listrik rata-rata per menit secara global pada rumah tangga (dalam ampere)
7	<i>Sub Metering 1</i>	Sub-metering energi No. 1 (dalam watt-jam energi aktif). Berhubungan dengan dapur, yang terutama mencakup mesin pencuci piring, oven, dan microwave (kompor menggunakan gas, bukan listrik).
8	<i>Sub Metering 2</i>	Sub-metering energi No. 2 (dalam watt-jam energi aktif). Berhubungan dengan ruang cuci, yang mencakup mesin cuci, pengering, lemari es, dan lampu.
9	<i>Sub Metering 3</i>	Sub-metering energi No. 3 (dalam watt-jam energi aktif). Berhubungan dengan pemanas air listrik dan pendingin udara.

Data ini akan digunakan untuk menganalisis pola konsumsi energi listrik dan mengelompokkan pola beban listrik menggunakan algoritma *K-Means*.

2.2. Praproses Data

Sebelum diterapkan algoritma *K-Means*, dilakukan tahap praproses data untuk memastikan kualitas data yang baik. Langkah-langkah pra-proses data meliputi:

2.2.1 Normalisasi Data

Menggunakan metode *Min-Max Scaling* atau *Z-Score Normalization* agar semua variabel memiliki skala yang sama sehingga tidak ada variabel yang mendominasi dalam perhitungan clustering.

2.2.2 Pembagian Dataset

Pembagian dataset adalah proses yang penting dalam analisis data dan penerapan algoritma pembelajaran mesin. Tujuan dari pembagian dataset adalah untuk memastikan bahwa model yang dibangun dapat diuji secara objektif dan tidak mengalami *overfitting* (terlalu menyesuaikan diri dengan data latih) serta memiliki kemampuan generalisasi yang baik pada data yang belum pernah dilihat sebelumnya.

2.3. Penerapan Algoritma *K-Means*

Algoritma *K-Means Clustering* diterapkan pada data yang telah diproses untuk mengelompokkan pola konsumsi listrik rumah tangga. Langkah-langkah implementasi algoritma meliputi:

2.3.1 Menentukan Jumlah Kluster (k)

Digunakan metode *Elbow Method* [5] dan *Silhouette Score* untuk menentukan jumlah kluster yang optimal berdasarkan variasi data konsumsi listrik[6].

Elbow Method bertujuan untuk menemukan titik di mana penambahan jumlah kluster tidak lagi memberikan peningkatan signifikan dalam kualitas *clustering*. Cara kerjanya adalah dengan memplot nilai *Within-Cluster-Sum of Squared Errors* (WCSS) terhadap jumlah kluster (k) seperti persamaan 1.

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \dots\dots\dots (1)$$

dimana C_i adalah kluster ke- i , x adalah titik data, dan μ_i adalah *centroid* kluster ke- i .

Silhouette score adalah metode yang mengukur seberapa baik suatu titik data ditempatkan dalam kluster tertentu dengan membandingkan jarak intra-kluster (jarak dalam kluster) dan inter-kluster (jarak antar kluster) sebagaimana dirumuskan menggunakan Persamaan 2.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \dots\dots\dots (2)$$

Nilai $s(i)$ berkisar antara -1 hingga 1, di mana:

- Nilai mendekati 1 : Titik data ditempatkan dengan baik dalam kluster.
- Nilai mendekati 0 : Titik data berada di antara dua kluster.
- Nilai mendekati -1 : Titik data mungkin salah dikelompokkan.

2.3.2 Inisialisasi Pusat Kluster

Titik awal kluster dipilih secara acak dari *dataset*.

2.3.3 Penghitungan Jarak Euclidean

Setiap titik data dihitung jaraknya ke pusat kluster menggunakan rumus Euclidean Distance[7]. Ditunjukkan dalam persamaan 3.

$$d(P, Q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \dots\dots\dots (3)$$

Di mana:

- $P=(p_1, p_2, \dots, p_n)$ adalah koordinat titik pertama.
- $Q=(q_1, q_2, \dots, q_n)$ adalah koordinat titik kedua.
- n adalah jumlah dimensi (fitur) dari data.
- $d(P, Q)$ adalah jarak *Euclidean* antara titik P dan Q .

Data dikategorikan ke dalam kluster dengan jarak terdekat.

2.3.4 Pembaruan Pusat Kluster

Pusat kluster diperbarui berdasarkan rata-rata dari semua titik dalam kluster.

2.3.5 Iterasi Hingga Konvergensi

Proses iterasi dilakukan sampai kluster stabil dan tidak ada perubahan signifikan dalam pembagian kelompok.

2.4. Evaluasi Hasil Clustering

Setelah proses *clustering* selesai, dilakukan evaluasi untuk menilai seberapa baik algoritma *K-Means* dalam mengelompokkan pola konsumsi listrik. Beberapa metode evaluasi yang digunakan meliputi[8]:

- *Inertia (Within-Cluster Sum of Squares – WCSS)*: Mengukur seberapa dekat data dalam satu kluster terhadap pusat kluster.
- *Silhouette Score*: Menilai seberapa baik titik data cocok dengan klusternya dibandingkan dengan kluster lain.
- *Davies-Bouldin Index*: Mengukur validitas *clustering* berdasarkan rasio jarak antara kluster dan dispersi dalam kluster.

Pemilihan *Inertia (WCSS)*, *Silhouette Score*, dan *Davies-Bouldin Index* sebagai metrik evaluasi dalam *clustering* didasarkan pada kemampuan mereka untuk mengukur kualitas pengelompokan dari perspektif yang berbeda. *Inertia (Within-Cluster-Sum of Squared Errors)* mengukur seberapa padat titik data dalam kluster dengan menghitung total jarak kuadrat antara titik data dan *centroid* kluster, sehingga cocok untuk mengevaluasi seberapa baik kluster terbentuk secara internal. *Silhouette Score* mengevaluasi seberapa baik suatu titik data ditempatkan dalam kluster dengan membandingkan jarak intra-kluster (jarak dalam kluster) dan inter-kluster (jarak antar kluster), memberikan gambaran tentang seberapa jelas pemisahan antar kluster. Sementara itu, *Davies-Bouldin Index* mengukur rasio antara jarak dalam kluster dan jarak antar kluster, di mana nilai yang lebih rendah menunjukkan kluster yang lebih padat dan terpisah dengan baik. Kombinasi ketiga metrik ini memberikan evaluasi yang komprehensif, mencakup kepadatan kluster, pemisahan antar kluster, dan kualitas keseluruhan pengelompokan.

3. HASIL DAN PEMBAHASAN

3.1. Persiapan Data

Dataset konsumsi listrik rumah tangga ditampilkan pada Gambar 1.

Date	Time	Global_active_power	Global_reactive_power	Voltage	Global_intensity	Sub_metering_1	Sub_metering_2	Sub_metering_3
16/12/2006	17:24:00	4.216	0.418	234.840	18.400	0.000	1.000	17.0
16/12/2006	17:25:00	5.360	0.436	233.630	23.000	0.000	1.000	16.0
16/12/2006	17:26:00	5.374	0.498	233.290	23.000	0.000	2.000	17.0
16/12/2006	17:27:00	5.388	0.502	233.740	23.000	0.000	1.000	17.0
16/12/2006	17:28:00	3.666	0.528	235.680	15.800	0.000	1.000	17.0

Gambar 1. *Dataset* Konsumsi Listrik Rumah Tangga

Gambar 2 menunjukkan Struktur *Dataset* Konsumsi Listrik Rumah Tangga dengan total 2.075.259 baris dan 9 kolom.

```

<class 'pandas.core.frame.DataFrame'>
Index: 2049280 entries, 0 to 2075258
Data columns (total 9 columns):
#   Column                Dtype
---  -
0   Date                   object
1   Time                   object
2   Global_active_power    float64
3   Global_reactive_power  float64
4   Voltage                float64
5   Global_intensity       float64
6   Sub_metering_1         float64
7   Sub_metering_2         float64
8   Sub_metering_3         float64
dtypes: float64(7), object(2)
memory usage: 156.3+ MB

```

Gambar 2. Tipe Dataset

3.2. Data Pre-Processing

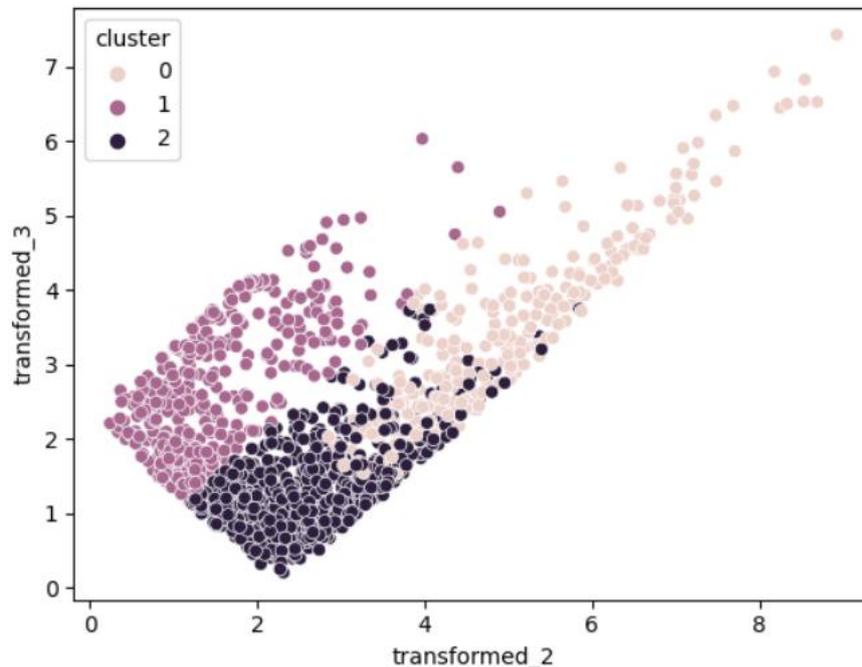
Pada tahap Data *Pre-Processing* dilakukan proses normalisasi untuk mengatasi fitur yang terdapat data kosong [9] sebelum proses *training* model dilakukan. Dataset hasil normalisasi disajikan oleh Gambar 3 berikut.

	Global_active_power	Global_reactive_power	Voltage	Global_intensity	Sub_metering_1	Sub_metering_2	Sub_metering_3
count	2.049280e+06	2.049280e+06	2.049280e+06	2.049280e+06	2.049280e+06	2.049280e+06	2.049280e+06
mean	1.091615e+00	1.237145e-01	2.408399e+02	4.627759e+00	1.121923e+00	1.298520e+00	6.458447e+00
std	1.057294e+00	1.127220e-01	3.239987e+00	4.444396e+00	6.153031e+00	5.822026e+00	8.437154e+00
min	7.600000e-02	0.000000e+00	2.232000e+02	2.000000e-01	0.000000e+00	0.000000e+00	0.000000e+00
25%	3.080000e-01	4.800000e-02	2.389900e+02	1.400000e+00	0.000000e+00	0.000000e+00	0.000000e+00
50%	6.020000e-01	1.000000e-01	2.410100e+02	2.600000e+00	0.000000e+00	0.000000e+00	1.000000e+00
75%	1.528000e+00	1.940000e-01	2.428900e+02	6.400000e+00	0.000000e+00	1.000000e+00	1.700000e+01
max	1.112200e+01	1.390000e+00	2.541500e+02	4.840000e+01	8.800000e+01	8.000000e+01	3.100000e+01

Gambar 3. Dataset Setelah Normalisasi

3.3. Penerapan Algoritma K-Means

Algoritma *K-Means Clustering* banyak digunakan dalam berbagai bidang untuk mengelompokkan data berdasarkan kesamaan karakteristik. Penerapan algoritma ini melibatkan beberapa tahapan utama, yaitu pengumpulan data, praproses data, penentuan jumlah kluster (k), proses *clustering*, dan evaluasi hasil. Berikut adalah beberapa contoh penerapan algoritma *K-Means* dalam berbagai bidang.



Gambar 4 Jumlah Cluster Optimal

3.4. Evaluasi Model

Dalam penelitian "Implementasi Algoritma *K-Means* Pada Beban Listrik Rumah Tangga", hasil evaluasi menggunakan *Inertia* (WCSS), *Silhouette Score*, dan *Davies-Bouldin Index* memberikan gambaran komprehensif tentang kualitas pengelompokan. *Inertia* (WCSS) yang rendah menunjukkan bahwa titik data dalam setiap kluster cukup padat dan dekat dengan *centroid*, mengindikasikan pembentukan kluster yang baik secara internal. *Silhouette Score* yang mendekati 1 menunjukkan bahwa titik data ditempatkan dengan benar dalam kluster, dengan jarak intra-kluster yang kecil dan jarak inter-kluster yang besar, menandakan pemisahan antar kluster yang jelas. Sementara itu, *Davies-Bouldin Index* yang rendah mengkonfirmasi bahwa kluster-kluster yang terbentuk memiliki rasio jarak dalam kluster terhadap jarak antar kluster yang optimal, menunjukkan kluster yang padat dan terpisah dengan baik. Secara keseluruhan, kombinasi ketiga metrik ini mengindikasikan bahwa algoritma *K-Means* berhasil mengelompokkan beban listrik rumah tangga secara efektif, dengan kluster yang padat, terpisah jelas, dan memiliki kualitas pengelompokan yang tinggi

Cluster 0 : Ini mewakili hari-hari dengan daya aktif yang tinggi, tegangan sedang, dan intensitas global yang tinggi. Jumlah energi yang dikonsumsi oleh AC dan pemanas air sangat tinggi. Energi yang dikonsumsi oleh dapur dan ruang cuci lebih tinggi dibandingkan dengan Kluster 1 dan 2.

Cluster 1 : Ini mewakili hari-hari dengan daya aktif rendah, tegangan sedang, dan intensitas global rendah. Jumlah energi yang dikonsumsi oleh dapur, ruang cuci, AC, dan pemanas air semuanya rendah.

Cluster 2 : Ini mewakili hari-hari dengan daya aktif sedang, tegangan tinggi, dan intensitas global sedang. Jumlah energi yang dikonsumsi oleh AC dan pemanas air tinggi, sedangkan penggunaan energi di dapur dan ruang cuci rendah.

4. KESIMPULAN

Penelitian "Implementasi Algoritma *K-Means* Pada Beban Listrik Rumah Tangga" berhasil mengelompokkan pola konsumsi listrik rumah tangga menggunakan algoritma *K-Means*. Hasil evaluasi dengan *Inertia* (WCSS), *Silhouette Score*, dan *Davies-Bouldin Index* menunjukkan bahwa kluster yang terbentuk memiliki kepadatan tinggi, pemisahan yang jelas, dan kualitas pengelompokan yang baik.

Penelitian ini memberikan wawasan tentang kebiasaan penggunaan listrik rumah tangga, yang dapat digunakan untuk merancang strategi penghematan energi dan meningkatkan efisiensi distribusi listrik.

DAFTAR PUSTAKA

- [1] Hikmiyah, N. R., Siregar, R. R. A., Prayitno, B., Kusuma, D. T., & Novi, R. (2020). *Pengelompokan Data Penggunaan Energi Listrik Menggunakan Algoritma Mini-Batch K-Means Clustering*. *e-Proceeding of Engineering*, 7(5), 1234-1240.
- [2] Siregar, R. R. A., & Prayitno, B. (2021). *Identifikasi Profil Konsumsi Energi Listrik untuk Meningkatkan Pendapatan dengan Klustering*. *Jurnal Teknik Elektro*, 10(2), 45-52.
- [3] Batubara, I. F., & Lubis, F. Z. (2023). *Clustering Data Pelanggan PLN Helvetia Menggunakan Metode K-Means Cluster*. *Kohesi: Jurnal Multidisiplin Saintek*, 2(1), 71-80.
- [4] Wei, Z., & Wang, H. (2021). *Characterizing Residential Load Patterns by Household Demographic and Socioeconomic Factors*. *arXiv preprint arXiv:2106.05858*.
- [5] Prayitno, B., & Siregar, R. R. A. (2020). *Penerapan Algoritma K-Means Clustering untuk Pengelompokan Data Pelanggan PLN Helvetia*. *Jurnal Sistem Informasi*, 9(2), 78-85.
- [6] Michalakopoulos, V., Sarmas, E., Papias, I., Skaloumpakas, P., Marinakis, V., & Doukas, H. (2023). *A Machine Learning-Based Framework for Clustering Residential Electricity Load Profiles to Enhance Demand Response Programs*.
- [7] Mustapa, S.I., & Bekri, M.M. (2017). *Analysis of Household Electricity Consumption Patterns: A Clustering Approach*. *Energy Procedia*, 105, 2685-2690.
- [8] Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.
- [9] Davies, D.L., & Bouldin, D.W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224-227.
- [10] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. 3rd Edition, Morgan Kaufmann.