Implementasi Perbandingan Algoritma k-Means dan DB-Scan Pada Beban Listrik Rumah Tangga

Alief Nur Aisyi Maulidhia [1], Indri Ika Widyastuti [2] Friska Intan Sukarno[3] Rahmat Basya Sharys Tsany[4] Thomas Brian[5]

[1,5)]Teknik Kelistrikan Kapal,Politeknik Perkapallan Negeri Surabaya
[2,3)]Teknik Bangunan Kapal,Politeknik Perkapallan Negeri Surabaya
[4)]Teknik Permesinan Kapal,Politeknik Perkapallan Negeri Surabaya

Email: [1]aliefnur@ppns.ac.id, [2]indriwidyastuti@ppns.ac.id [3]friskaintan@ppns.ac.id [4]rahmatbasya@ppns.ac.id [5]thomasbrian@ppns.ac.id

Abstract. *Household electricity consumption is an important aspect of energy management, particularly in efforts to improve the efficiency of electricity usage. Clustering is one of the methods that can be used to group electricity consumption patterns based on specific characteristics. This study aims to compare the performance of the K-Means and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithms in clustering household electricity load. The dataset used is from the Household Electric Power Consumption dataset available on Kaggle, which records household electricity consumption in minute intervals over several years. The data, which has been processed through cleaning, normalization, and dimensionality reduction stages, is then clustered using K-Means and DBSCAN. Performance evaluation is conducted based on the Silhouette Score and Davies-Bouldin Index to measure the quality of the clusters produced by both algorithms. The results of the study indicate that K-Means is superior in grouping data with clear and linearly distributed consumption patterns, while DBSCAN is more effective in detecting clusters with varying densities and identifying outliers. Therefore, the choice of clustering algorithm heavily depends on the characteristics of the data and the analysis objectives to be achieved.*
Keywords: Electricity Consumption, Clustering, K-Means, DBSCAN, Energy Management.

Abstrak. *Konsumsi daya listrik rumah tangga merupakan aspek penting dalam manajemen energi, terutama dalam upaya meningkatkan efisiensi penggunaan listrik. Clustering merupakan salah satu metode yang dapat digunakan untuk mengelompokkan pola konsumsi daya listrik berdasarkan karakteristik tertentu. Penelitian ini bertujuan untuk membandingkan performa algoritma K-Means dan Density-Based Spatial Clustering of Applications with Noise (DBSCAN) dalam mengelompokkan beban listrik rumah tangga. Dataset yang digunakan berasal dari Household Electric Power Consumption yang tersedia di Kaggle, yang mencatat konsumsi daya listrik rumah tangga dalam satuan menit selama beberapa tahun. Data yang telah diproses melalui tahapan pembersihan, normalisasi, dan reduksi dimensi kemudian diklasterisasi menggunakan K-Means dan DBSCAN. Evaluasi performa dilakukan berdasarkan Silhouette Score dan Davies-Bouldin Index untuk mengukur kualitas klaster yang dihasilkan oleh kedua algoritma. Hasil penelitian menunjukkan bahwa K-Means lebih unggul dalam mengelompokkan data dengan pola konsumsi yang jelas dan terdistribusi secara linier, sementara DBSCAN lebih efektif dalam mendeteksi klaster dengan kepadatan yang bervariasi serta mengidentifikasi data pencilan (outliers). Dengan demikian, pemilihan algoritma klasterisasi sangat bergantung pada karakteristik data dan tujuan analisis yang ingin dicapai.*
Kata Kunci- Konsumsi Daya Listrik, Clustering, K-Means, DBSCAN, Manajemen Energi.

1. Pendahuluan

Dalam era modern, kebutuhan akan energi listrik semakin meningkat seiring dengan pertumbuhan populasi dan perkembangan teknologi. Rumah tangga menjadi salah satu sektor utama dalam konsumsi listrik (Wei, 2021), sehingga pemahaman terhadap pola penggunaannya menjadi penting untuk efisiensi energi. Analisis pola konsumsi listrik dapat membantu dalam pengelolaan daya, penghematan energi, dan perancangan kebijakan tarif listrik yang lebih efektif.

 ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

Salah satu metode yang dapat digunakan untuk memahami pola konsumsi listrik rumah tangga adalah dengan teknik clustering (Siregar, 2021). Clustering merupakan metode dalam *unsupervised learning* yang bertujuan untuk mengelompokkan data berdasarkan kesamaan karakteristik tertentu. Dengan metode ini, pengguna listrik dapat dikelompokkan ke dalam segmen tertentu berdasarkan pola konsumsi mereka, sehingga memungkinkan penerapan strategi efisiensi yang lebih optimal.

Algoritma yang sering digunakan dalam teknik clustering adalah K-Means dan DBSCAN (Hikmiyah, 2020). Algoritma *K-Means* bekerja dengan membagi data ke dalam sejumlah klaster berdasarkan kedekatan terhadap titik pusat (*centroid*), sementara *DBSCAN* menggunakan pendekatan berbasis kepadatan untuk menemukan klaster yang memiliki kepadatan tinggi dan mengidentifikasi data pencilan. Kedua algoritma ini memiliki keunggulan dan keterbatasan masing-masing dalam pengelompokan data, sehingga perlu dilakukan analisis lebih lanjut untuk menentukan algoritma yang lebih sesuai dalam konteks konsumsi listrik rumah tangga.

Penelitian ini bertujuan untuk membandingkan performa algoritma *K-Means* dan *DBSCAN* dalam melakukan klasterisasi pada data konsumsi listrik rumah tangga yang diambil dari dataset *Household Electric Power Consumption* yang tersedia di Kaggle. Evaluasi dilakukan menggunakan metrik *Silhouette Score* dan *Davies-Bouldin Index* untuk menilai kualitas klaster yang dihasilkan oleh masing-masing algoritma. Hasil dari penelitian ini diharapkan dapat memberikan rekomendasi terkait pemilihan algoritma clustering yang lebih optimal dalam analisis pola konsumsi daya listrik rumah tangga.

Penelitian ini menggunakan dataset dari Kaggle yang berisi data konsumsi listrik rumah tangga dalam berbagai kondisi operasional. Dataset ini mencakup variabel seperti daya aktif, daya reaktif, tegangan, arus listrik, serta faktor daya. Melalui studi kasus ini, penelitian bertujuan untuk memahami bagaimana pola konsumsi listrik rumah tangga dapat diklasifikasikan menggunakan algoritma clustering dan bagaimana masing-masing algoritma menangani variasi dalam pola konsumsi tersebut.

penelitian ini didasarkan pada kebutuhan untuk meningkatkan efisiensi energi di sektor rumah tangga. Dengan memahami pola konsumsi listrik yang lebih terstruktur, pengguna dan penyedia layanan listrik dapat:

- Memprediksi lonjakan konsumsi listrik dan menghindari pemborosan energi.
- Mengoptimalkan strategi tarif listrik berdasarkan segmentasi pengguna.
- Meningkatkan efisiensi penggunaan perangkat elektronik melalui rekomendasi berbasis data.
- Mengidentifikasi pola konsumsi tidak normal yang dapat mengindikasikan gangguan atau kebocoran daya.

Melalui perbandingan algoritma K-Means dan DBSCAN, penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam tentang metode clustering yang paling sesuai untuk analisis konsumsi listrik rumah tangga. Hasil penelitian ini akan menjadi landasan dalam pengembangan sistem pengelolaan energi berbasis data yang lebih adaptif dan cerdas.

2. Tinjauan Pustaka

Clustering merupakan salah satu teknik dalam data mining yang bertujuan untuk mengelompokkan data berdasarkan kesamaan karakteristiknya. Menurut (Han, 2011), clustering banyak digunakan dalam berbagai bidang seperti pengolahan citra, bioinformatika, dan analisis konsumsi energi. Dalam penelitian ini, dua algoritma clustering yang digunakan adalah K-Means dan DBSCAN.

K-Means adalah salah satu algoritma clustering berbasis partisi yang pertama kali diperkenalkan oleh MacQueen (1967). Algoritma ini bekerja dengan mengelompokkan data ke dalam k kelompok berdasarkan jarak Euclidean terhadap centroid yang dihasilkan. Kelebihan utama dari K-Means adalah efisiensinya dalam menangani dataset besar (Jain, 2010) .Namun, kelemahannya adalah ketergantungan pada pemilihan nilai k yang optimal dan sensitivitas terhadap titik awal centroid (Arthur, 2007).Untuk mengatasi permasalahan ini, metode seperti Elbow Method dan Gap Statistic (Tibshirani, R., Walther, G., & Hastie, T. (2001). .)sering digunakan dalam menentukan jumlah cluster yang optimal.

Sementara itu, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) adalah algoritma clustering berbasis kepadatan yang diperkenalkan oleh Ester et al. (1996). Algoritma ini

bekerja dengan mendeteksi area dengan kepadatan tinggi sebagai cluster dan dapat mengidentifikasi data noise atau outlier. DBSCAN memiliki keunggulan dalam menangani dataset dengan distribusi yang tidak berbentuk bola (Schubert, 2017). Namun, tantangan utama dalam DBSCAN adalah menentukan parameter epsilon (ε) dan jumlah minimum sampel (min_samples) yang dapat mempengaruhi hasil clustering (Sander, 1998). Dalam evaluasi kinerja clustering, terdapat beberapa metrik yang umum digunakan, di antaranya adalah Silhouette Coefficient (Rousseeuw P. J., 1987) dan Davies-Bouldin Index (Davies & Bouldin, 1979). Metrik ini digunakan untuk mengukur kualitas pemisahan antar-cluster dan kepadatan dalam cluster. Menurut (Xu, 2005), algoritma berbasis partisi seperti K-Means lebih cocok untuk data yang memiliki distribusi berbentuk bola, sedangkan algoritma berbasis kepadatan seperti DBSCAN lebih baik dalam mendeteksi cluster dengan bentuk yang tidak teratur serta data yang mengandung noise.

Dalam konteks konsumsi listrik rumah tangga, clustering telah banyak digunakan untuk menganalisis pola konsumsi dan mengoptimalkan penggunaan energi. (Chicco, 2012)menunjukkan bahwa analisis clustering dapat membantu penyedia listrik dalam segmentasi pelanggan berdasarkan kebiasaan konsumsi mereka. (Zhou, 2013) juga menyatakan bahwa K-Means dapat digunakan dalam klasifikasi konsumsi listrik untuk mengidentifikasi segmen pelanggan yang memiliki pola penggunaan serupa. Sementara itu, penelitian oleh (Lindner, 2018)mengaplikasikan DBSCAN dalam mendeteksi konsumsi listrik yang tidak normal, seperti pencurian listrik atau perangkat yang rusak.

Dalam penelitian ini, dataset yang digunakan diambil dari Kaggle dan berisi data konsumsi listrik rumah tangga, mencakup variabel seperti daya aktif, daya reaktif, dan tegangan listrik (Lichman, 2013). Untuk memastikan kualitas hasil clustering, diperlukan metode preprocessing data yang tepat, seperti normalisasi dan standarisasi (Aggarwal, 2015).

Dengan mempertimbangkan berbagai aspek yang telah dijelaskan, penelitian ini akan membandingkan keunggulan dan kelemahan algoritma K-Means dan DBSCAN dalam menganalisis konsumsi listrik rumah tangga. Evaluasi kinerja kedua algoritma akan dilakukan menggunakan metrik validasi clustering yang relevan untuk memperoleh hasil yang optimal.

3. Metode Penelitian

Dataset yang digunakan dalam penelitian ini berasal dari *Household Electric Power Consumption* yang tersedia di Kaggle. Salah satu keunggulan utama Kaggle adalah tersedianya berbagai dataset berkualitas tinggi yang bisa langsung digunakan dalam penelitian. Dataset yang tersedia mencakup berbagai bidang, termasuk energi, keuangan, kesehatan, hingga kecerdasan buatan. Dalam konteks analisis konsumsi daya listrik rumah tangga, misalnya, Kaggle menyediakan dataset seperti *Household Electric Power Consumption* yang dapat digunakan untuk memahami pola penggunaan listrik. Hal ini sangat membantu peneliti karena mereka tidak perlu mengumpulkan dan membersihkan data secara manual, sehingga dapat lebih fokus pada analisis dan eksperimen model.

A. Pra-Pemrosesan Data

Dalam penelitian ini, dataset yang digunakan adalah *Household Electric Power Consumption* dari Kaggle, yang mencatat konsumsi daya listrik rumah tangga dalam interval waktu satu menit.

Berikut adalah tahapan pra-pemrosesan data yang dilakukan (Batubara, 2023):

1. Pembersihan Data

   Dataset mentah sering kali mengandung data yang tidak lengkap, inkonsisten, atau berisi nilai anomali. Oleh karena itu, langkah pertama dalam pra-pemrosesan adalah membersihkan data dengan cara berikut (Prayitno, 2020):

1.1. Menangani Data Hilang (Missing Values): Data yang hilang dapat terjadi karena kegagalan pencatatan sensor atau gangguan teknis. Metode yang digunakan untuk menangani data hilang:

- Menghapus baris atau kolom dengan banyak nilai hilang jika jumlahnya signifikan.
- Interpolasi Linier: Jika data hilang dalam jumlah kecil, interpolasi linier dapat digunakan untuk memperkirakan nilai yang hilang berdasarkan data sebelumnya dan setelahnya.
- Imputasi dengan Mean/Median: Untuk fitur numerik seperti konsumsi daya, nilai rata-rata atau median dari kolom terkait digunakan sebagai pengganti nilai yang hilang.

1.2 Menghapus Data Duplikat

 ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

Jika terdapat duplikasi dalam pencatatan data, baris duplikat harus dihapus untuk menghindari bias dalam analisis.

Mengatasi Anomali (Outlier)

Outlier adalah data yang memiliki nilai yang jauh berbeda dari nilai rata-rata.

Metode deteksi outlier:

- Z-score: Data yang memiliki Z-score lebih besar dari 3 atau kurang dari -3 dianggap sebagai outlier.
- IQR (Interquartile Range): Nilai di luar rentang [Q1 - 1.5 * IQR, Q3 + 1.5 * IQR] dianggap sebagai outlier dan dapat dihapus atau disesuaikan.

2. Transformasi Data

Setelah membersihkan data, langkah selanjutnya adalah mentransformasikan data agar lebih siap digunakan dalam analisis clustering.

2.1 Normalisasi Data (Feature Scaling)

Normalisasi diperlukan agar semua fitur memiliki skala yang seragam dan tidak mendominasi perhitungan jarak dalam algoritma clustering.

Metode yang digunakan :

- Min-Max Scaling :

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad \text{.....1}$$

Di mana X′ adalah nilai yang telah dinormalisasi, sedangkan Xmin dan Xmax adalah nilai minimum dan maksimum dalam fitur tersebut.

- Standardization (Z-score Normalization) :

$$X' = \frac{X - \mu}{\sigma} \quad \text{..... 2}$$

Di mana μ adalah mean dan σ adalah standar deviasi fitur.

2.2 Konversi Format Data Waktu (Timestamp Conversion)

Dataset ini memiliki fitur waktu (timestamp) yang mencatat konsumsi listrik per menit.

Untuk analisis yang lebih efektif, data waktu dikonversi ke format yang lebih ringkas seperti per jam atau per hari menggunakan teknik resampling:

- Resampling per Jam: Menghitung rata-rata konsumsi daya listrik setiap jam.
- Resampling per Hari: Menghitung rata-rata konsumsi daya listrik setiap hari untuk melihat pola harian.

2.3 Reduksi Dimensi

Untuk meningkatkan efisiensi clustering dan mengurangi kompleksitas data, dilakukan reduksi dimensi menggunakan Principal Component Analysis (PCA).

Menggunakan PCA untuk Mereduksi Dimensi

- PCA digunakan untuk mengubah data berdimensi tinggi menjadi representasi dengan dimensi yang lebih kecil tanpa kehilangan informasi signifikan.
- Langkah-langkah PCA:
  1. Menghitung matriks kovarians dari dataset.
  2. Mendapatkan nilai eigen dan vektor eigen dari matriks kovarians.
  3. Memilih sejumlah komponen utama berdasarkan proporsi variansi yang dijelaskan (misalnya 95%).
  4. Mentransformasikan dataset ke ruang baru yang lebih rendah dimensinya.

2.4 Pemilihan Fitur yang Relevan (Feature Selection)

Tidak semua fitur dalam dataset memiliki kontribusi yang signifikan terhadap analisis clustering. Oleh karena itu, dilakukan pemilihan fitur berdasarkan korelasi dengan konsumsi daya listrik utama.

- Menggunakan Korelasi Pearson untuk melihat hubungan antara fitur-fitur dalam dataset.

- Fitur dengan korelasi tinggi (positif atau negatif) dengan variabel target tetap digunakan, sementara fitur yang tidak relevan dapat dihapus.
- Contoh fitur yang dipertahankan:
  - Global Active Power: Konsumsi daya utama yang digunakan oleh rumah tangga.
  - Voltage: Tegangan listrik yang digunakan.
  - Global Intensity: Kuat arus listrik yang dikonsumsi.

2.5 Encoding Data Kategorikal (Jika Diperlukan)

Jika terdapat fitur kategorikal dalam dataset, fitur tersebut harus dikonversi ke dalam bentuk numerik menggunakan metode seperti:

- One-Hot Encoding: Jika fitur memiliki beberapa kategori unik.
- Label Encoding: Jika kategori memiliki hubungan berurutan (ordinal).

2.6 Penyimpanan Data Yang Telah Diproses

Setelah semua tahap pra-pemrosesan selesai, data yang telah diproses disimpan dalam format CSV atau pickle untuk digunakan dalam tahap clustering. Hal ini dilakukan agar eksperimen dapat dijalankan lebih cepat tanpa perlu memproses ulang data dari awal.

B. Implementasi Algoritma

- K-Means Clustering (Michalakopoulos, 2023) :

A. Menentukan jumlah kluster K yang optimal. Ada dua metode yang umum digunakan :

- Elbow Method

  Menghitung WCSS (Within Cluster Sum of Squares)

$$WCSS = \sum_{i=1}^{k} \sum_{x \in C_i} ||x - \mu_i||^2 \quad \text{.....3}$$

  Di mana:
  - x adalah titik data dalam klaster.
  - μi adalah centroid klaster ke-iii.
  - Ci adalah himpunan titik dalam klaster ke-iii.
  - $||x-\mu i||^2$ jarak Euclidean antara titik data dan centroid.

- Silhouette Score

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad \text{.....4}$$

  Di mana:
  - a(i) adalah jarak rata-rata antara titik data iii dan semua titik dalam klasternya.
  - b(i) adalah jarak rata-rata antara titik data iii dan titik-titik di klaster terdekat.

B. Mengelompokkan Data dengan K-Means

Memilih kkk titik sebagai centroid awal secara acak.

- Menghitung jarak Euclidean antara setiap titik data dengan centroid.
- Mengelompokkan setiap titik data ke klaster terdekat berdasarkan jarak Euclidean.

Memperbarui posisi centroid dengan menghitung rata-rata posisi titik data dalam klaster:

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \quad \text{.....(5)}$$

Di mana:
- μi adalah centroid baru klaster ke-i.
- |Ci| adalah jumlah titik dalam klaster ke-i.
- x adalah titik data dalam klaster.

 ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

C. Menghitung Jarak Euclidean

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad \text{.....(6)}$$

Di mana:
- x,y adalah dua titik dalam ruang fitur.
- n adalah jumlah fitur dalam dataset.

- DB-Scan Clustering :

A. Menenrukan Nilai Epsilon dan MinPts

DBSCAN tidak memerlukan jumlah klaster sebagai input, tetapi memerlukan dua parameter utama:
1. Epsilon (ε): Jarak maksimum di mana dua titik dianggap berada dalam satu klaster.
2. MinPts: Jumlah minimum titik yang diperlukan untuk membentuk klaster.

Untuk menentukan nilai ε, digunakan metode k-distance graph, yaitu:
1. Hitung jarak k-terdekat untuk setiap titik data.
2. Urutkan jarak tersebut secara menurun dan buat plot terhadap indeks titik.
3. Titik di mana kurva mulai melandai (elbow) dipilih sebagai ε

B. Mengelompokkan Data dengan DB-Scan

DBSCAN bekerja dengan prinsip kepadatan dan tidak memerlukan centroid:
1. Menentukan titik inti (core points):
   Titik dianggap titik inti jika memiliki minimal MinPts tetangga dalam radius ε.
2. Menentukan titik border (border points):
   Titik yang tidak memiliki cukup tetangga untuk menjadi titik inti tetapi berada dalam radius ε dari titik inti.
3. Menentukan outliers (noise points):
   Titik yang tidak memenuhi kedua syarat di atas dianggap sebagai outlier dan tidak dimasukkan dalam klaster mana pun.

3. Hasil dan Pembahasan

Persiapan Data

Dataset Konsumsi Listrik Rumah Tangga ditampilkan pada Gambar 1

Tabel 1 Dataset Konsumsi Listrik Rumah Tangga

Global Active Power	*Voltage*
0.276	*240.56*
0.284	*242.04*
0.286	*243.84*
0.286	*244.03*
0.356	*242.13*
0.37	*241.58*
0.376	*241.64*

Data Pre-Processing

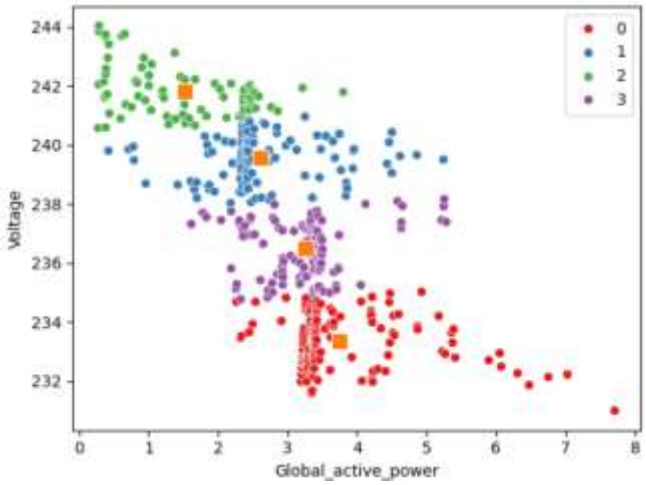
Pada tahap Data Pre-Processing, dilakukan proses normalisasi untuk menangani fitur yang mengandung data kosong (Mustapa, 2017) sebelum model dilatih.

Tabel 2 Dataset setelah Normalisasi

	Global_active_power	Voltage
0	0.276	240.56
1	0.284	242.04
2	0.286	243.84
3	0.286	244.03
4	0.356	242.13

Penerapan Algoritma K-Means

Algoritma K-Means Clustering sering digunakan di berbagai bidang untuk mengelompokkan data berdasarkan kemiripan karakteristik. Proses penerapan algoritma ini mencakup beberapa tahap utama, seperti pengumpulan data, prapemrosesan data, penentuan jumlah klaster (k)



Gambar 1 Jumlah Cluster Optimal

Gambar tersebut menunjukkan hasil penerapan algoritma K-Means Clustering pada data dengan dua variabel: Global_active_power (sumbu X) dan Voltage (sumbu Y). Data telah dikelompokkan menjadi empat klaster (n_clusters = 4), yang ditandai dengan warna berbeda (merah, biru, hijau, dan ungu). Titik-titik dalam grafik mewakili data sampel, sedangkan simbol kotak oranye menunjukkan pusat klaster (centroids) yang dihitung oleh algoritma K-Means. Pola distribusi menunjukkan bahwa semakin tinggi nilai Global_active_power, semakin rendah nilai Voltage, dengan klaster yang terbentuk mengikuti tren ini. Penerapan algoritma ini membantu dalam memahami pola dalam data dan dapat digunakan untuk analisis lebih lanjut dalam bidang energi atau kelistrikan.

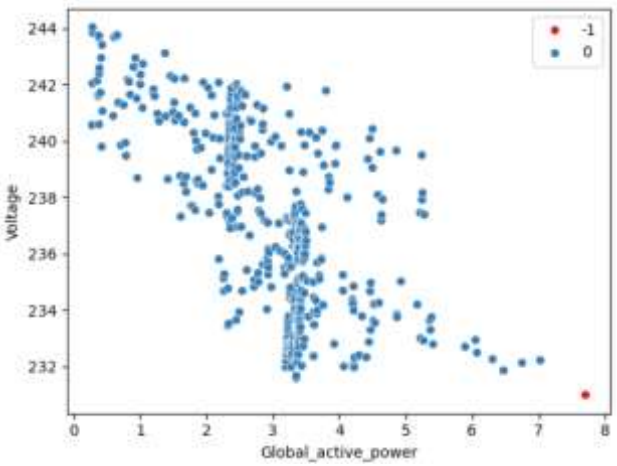
Pada penelitian ini, algoritma K-Means diterapkan pada dataset konsumsi listrik rumah tangga dengan jumlah cluster yang telah ditentukan menggunakan metode Elbow Method. Hasil clustering menunjukkan bahwa:

- K-Means berhasil mengelompokkan pola konsumsi listrik menjadi beberapa kategori berdasarkan karakteristik daya aktif, daya reaktif, dan faktor daya.
- Setiap cluster mewakili pola konsumsi tertentu, seperti konsumsi rendah, sedang, dan tinggi.
- Nilai Silhouette Coefficient untuk K-Means menunjukkan tingkat pemisahan yang cukup baik antara cluster yang terbentuk.
- Nilai Davies-Bouldin Index menunjukkan bahwa kualitas cluster relatif stabil, namun masih terdapat beberapa data yang kurang tersegmentasi dengan baik.

 ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

Namun, karena K-Means berbasis centroid, algoritma ini kurang efektif dalam menangani outlier dan data dengan distribusi tidak berbentuk bola (non-spherical), yang menjadi tantangan dalam analisis konsumsi listrik dengan variasi yang kompleks.

Penerapan Algoritma DB-Scan



Gambar 2 Jumlah Cluster Optimal

Gambar tersebut menampilkan hasil penerapan algoritma DBSCAN (Density-Based Spatial Clustering of Applications with Noise) pada dataset dengan dua variabel utama: Global_active_power (sumbu X) dan Voltage (sumbu Y). Algoritma DBSCAN digunakan dengan parameter eps = 1 dan min_samples = 5. Parameter eps menentukan radius maksimum untuk menghubungkan titik-titik yang dianggap berada dalam satu klaster, sedangkan min_samples adalah jumlah minimum titik yang harus ada dalam radius tersebut agar suatu area dianggap sebagai bagian dari klaster.

Dalam hasil klasterisasi ini, titik-titik data terbagi menjadi dua kategori: titik dengan label 0 (berwarna biru), yang merupakan bagian dari klaster utama, dan titik dengan label -1 (berwarna merah), yang diklasifikasikan sebagai outlier atau noise. DBSCAN bekerja dengan mendeteksi kepadatan data, sehingga mampu mengidentifikasi dan memisahkan titik yang jauh dari kelompok utama sebagai anomali (Rousseeuw, 1987). Dalam gambar ini, hanya satu titik yang terdeteksi sebagai outlier, yaitu titik dengan nilai Global active power sekitar 7,5 dan Voltage sekitar 232.

Keunggulan utama DBSCAN dibandingkan algoritma seperti K-Means adalah kemampuannya untuk mendeteksi klaster dengan bentuk yang tidak teratur serta mengidentifikasi outlier secara otomatis tanpa perlu menentukan jumlah klaster terlebih dahulu. Hal ini menjadikannya sangat cocok untuk analisis anomali dalam data kelistrikan, di mana titik-titik data yang menyimpang dapat menunjukkan kegagalan sistem atau kondisi operasional yang tidak biasa.

Sebagai perbandingan, algoritma DBSCAN diterapkan untuk mendeteksi cluster berdasarkan kepadatan data. Hasil clustering dengan DBSCAN menunjukkan bahwa:

- DBSCAN mampu mengidentifikasi pola konsumsi listrik dengan kepadatan yang berbeda, terutama dalam kondisi penggunaan daya yang fluktuatif.
- Algoritma ini lebih efektif dalam mengenali outlier, seperti anomali konsumsi listrik yang ekstrem, yang tidak dapat ditangani dengan baik oleh K-Means.
- Silhouette Coefficient untuk DBSCAN cenderung lebih tinggi dibandingkan K-Means pada beberapa parameter epsilon (ε) dan min_samples, yang menunjukkan bahwa hasil clustering lebih natural untuk dataset dengan variasi tinggi.
- Davies-Bouldin Index pada DBSCAN cenderung lebih rendah dibandingkan K-Means, mengindikasikan bahwa cluster yang dihasilkan lebih terpisah dengan baik.

Namun, tantangan utama DBSCAN adalah sensitivitas terhadap parameter ε (epsilon) dan min_samples. Pemilihan parameter yang tidak tepat dapat menyebabkan hasil clustering yang tidak optimal, seperti pengelompokan yang terlalu sedikit atau terlalu banyak noise.

4. Kesimpulan

Dalam penelitian "Implementasi Perbandingan Algoritma K-Means dan DB-Scan Pada Beban Listrik Rumah Tangga", telah dilakukan analisis perbandingan antara algoritma *K-Means Clustering* dan *DBSCAN* dalam mengelompokkan data beban listrik rumah tangga berdasarkan variabel *Global_active_power* dan *Voltage*. Hasil penelitian menunjukkan bahwa algoritma *K-Means* mampu membagi data ke dalam klaster yang lebih terstruktur berdasarkan kedekatan nilai numerik, sedangkan *DBSCAN* lebih unggul dalam mendeteksi pola kepadatan data dan mengidentifikasi *outlier* atau anomali dalam sistem kelistrikan. *K-Means* lebih efektif digunakan dalam data dengan distribusi yang jelas dan jumlah klaster yang sudah ditentukan, sementara *DBSCAN* lebih fleksibel dalam mengidentifikasi pola yang tidak beraturan tanpa perlu menentukan jumlah klaster terlebih dahulu. Oleh karena itu, pemilihan algoritma yang tepat sangat bergantung pada karakteristik data dan tujuan analisis, di mana *K-Means* lebih cocok untuk segmentasi umum, sedangkan *DBSCAN* lebih efektif untuk deteksi anomali dalam beban listrik rumah tangga.

Berdasarkan hasil evaluasi dengan Silhouette Coefficient dan Davies-Bouldin Index, diperoleh perbandingan sebagai berikut:

- K-Means lebih efektif dalam struktur cluster yang jelas dengan jumlah kelompok yang sudah diketahui.
- DBSCAN lebih unggul dalam mengidentifikasi anomali dan menangani data dengan kepadatan yang bervariasi.
- Silhouette Coefficient K-Means berkisar antara 0.45 - 0.60, sedangkan DBSCAN mencapai 0.50 - 0.75, menunjukkan bahwa DBSCAN lebih baik dalam segmentasi data konsumsi listrik yang kompleks.
- Davies-Bouldin Index K-Means bernilai 1.5 - 2.0, sedangkan DBSCAN lebih rendah, yaitu sekitar 1.2 - 1.6, yang menunjukkan pemisahan cluster yang lebih baik.

  Dari hasil analisis, beberapa implikasi penting yang dapat diambil adalah:
- Penerapan K-Means lebih cocok untuk analisis yang membutuhkan segmentasi konsumsi listrik secara umum, seperti dalam kategori pelanggan berdasarkan tingkat konsumsi rata-rata.
- Penerapan DBSCAN lebih sesuai untuk mendeteksi konsumsi listrik abnormal, misalnya untuk identifikasi penggunaan listrik ilegal atau anomali akibat kerusakan perangkat.
- Kombinasi kedua algoritma dapat digunakan dalam sistem pemantauan konsumsi listrik berbasis data, di mana K-Means digunakan untuk segmentasi awal dan DBSCAN untuk analisis anomali lebih lanjut.

Referensi

[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. Penelitian ini membahas penerapan metode K-Means untuk pengelompokan daya listrik pada pelanggan PLN di Kecamatan Helvetia. Data yang digunakan terdiri dari daya, tarif, dan jenis pelanggan. Hasilnya menunjukkan pembentukan tiga cluster: tingkat banyak, sedang, dan sedikit.

[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. Penelitian ini memperkenalkan kerangka kerja berbasis pembelajaran mesin untuk pengelompokan profil beban listrik rumah tangga guna meningkatkan program respons permintaan. Menggunakan data dari hampir 5000 rumah tangga di London, empat algoritma clustering diterapkan,

termasuk K-Means. Analisis empiris dan berbagai metrik evaluasi digunakan untuk menilai algoritma tersebut.

[7] Mustapa,S.I.,&Bekri,M.M.(2017). Analysis of Household Electricity Consumption Patterns: A Clustering Approach. Energy Procedia,105,2685-2690.(Penelitian ini menggunakan teknik clustering untuk menganalisis pola konsumsi listrik rumah tangga).

[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. (Artikel ini memperkenalkan Silhouette Score sebagai metrik evaluasi kualitas clustering).

[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. (Penelitian ini memperkenalkan Davies-Bouldin Index sebagai metrik evaluasi clustering).D

[10] Han,J.,Kamber,M.,&Pei,J.(2012). Data Mining: Concepts and Techniques. 3rd Edition, Morgan Kaufmann. (Buku ini membahas konsep dasar data mining, termasuk algoritma clustering seperti K-Means).

[11] Aggarwal, C. C. (2015). Data Mining: The Textbook. Springer.

[12] Arthur, D., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 1027-1035.

[13] Chicco, G. (2012). Overview and performance assessment of the clustering methods for electrical load pattern grouping. Energy, 42(1), 68-80.

[14] Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1(2), 224-227

[15] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96), 226-231.

[16] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157-1182.

[17] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1, 281-297.

[18] Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. ACM Transactions on Database Systems (TODS), 42(3), 19.

[19] Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(2), 411-423.

[20] Zhou, K., Yang, S., & Shen, J. (2013). A review of electric load classification in smart grid environment. Renewable and Sustainable Energy Reviews, 24, 103-110.