



## *Clustering Electricity Distribution Data Using Density-Based Spatial Clustering of Applications With Noise (DBSCAN) Algorithm*

### **Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Density Based Spatial Clustering of Application With Noise (DBSCAN)**

Miftah Farid<sup>1</sup>, Fitri Insani<sup>2\*</sup>, Surya Agustian<sup>3</sup>, Liza Afriyanti<sup>4</sup>

<sup>1,2,3,4</sup>Program Studi Teknik Informatika, Fakultas Sains dan Teknologi,  
Universitas Islam Negeri Sultan Syarif Kasim, Indonesia

E-Mail: <sup>1</sup>faridhimssi@gmail.com, <sup>2</sup>fitri.insani@uin-suska.ac.id,  
<sup>3</sup>surya.agustian@uin-suska.ac.id, <sup>4</sup>liza.afriyanti@uin-suska.ac.id

Received May 25th 2024; Revised Jun 20th 2024; Accepted Jun 21th 2024  
Corresponding Author: Fitri insani

#### **Abstract**

*In today's world, electricity has become an essential necessity in life, as most human activities depend on electricity. The demand for electricity varies significantly across different regions in Indonesia, influenced by various factors and the specific characteristics of each area. PLN, Indonesia's state-owned electricity company, publishes detailed electricity usage statistics from 2014 to 2022 for all 35 provinces. This data offers valuable insights for forecasting the demand, tracking historical trends to predict area development, prioritizing high-demand area for efficiency and energy conservation, and so on. One of the tools to achieve those goals is clustering the area for its characteristics and features localization. This research carried out a clustering process by dividing the data into three main regions: Sumatra, Java-Bali, and Kalimantan-Sulawesi, while the Papua region was not analyzed due to the limited number of regions/provinces. The method used is Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Parameter tuning is used to obtain optimal results based on the silhouette score. The clustering results can provide an overview of the unique profile of electricity consumption in each region, with the best silhouette score of 0.62 for the Java-Bali region, 0.67 for Kalimantan-Sulawesi, and 0.64 for Sumatra. This research shows that the DBSCAN algorithm can be used to group electricity distribution with effective results.*

*Keywords: Clustering, Data Analysis, DBSCAN, Electricity Distribution*

#### **Abstrak**

Pada masa kini, listrik sudah menjadi kebutuhan penting dalam kehidupan, karena kebanyakan aktivitas manusia bergantung pada listrik. Kebutuhan listrik pada setiap wilayah di Indonesia dipengaruhi oleh sejumlah faktor dan karakteristik khusus masing-masing. PLN mempublikasikan statistik penggunaan listrik untuk setiap wilayah di Indonesia dari tahun 2014 hingga 2022, yang terdiri dari 35 provinsi di Indonesia. Data ini menawarkan wawasan berharga untuk prediksi permintaan listrik, pelacakan tren historis untuk memprediksi pengembangan wilayah, memprioritaskan wilayah dengan permintaan tinggi untuk efisiensi dan konservasi energi, dan lain sebagainya. Salah satu alat untuk mencapai tujuan tersebut adalah dengan mengelompokkan (clustering) wilayah berdasarkan karakteristik dan ciri-ciri wilayahnya. Penelitian ini melakukan proses clustering dengan membagi data pada tiga regional utama: Sumatera, Jawa-Bali, dan Kalimantan-Sulawesi, sementara regional Papua tidak dianalisis karena jumlah wilayah/propinsi yang terbatas. Metode yang dipakai adalah *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN). Tuning parameter dengan cara pencarian *grid* dilakukan untuk memperoleh hasil optimal berdasarkan *silhouette score*. Hasil clustering dapat memberikan gambaran keunikan profil konsumsi listrik di tiap wilayah, dengan *silhouette score* terbaik sebesar 0.62 untuk regional Jawa-Bali, 0.67 untuk Kalimantan-Sulawesi, dan 0.64 untuk Sumatera. Penelitian menunjukkan bahwa algoritma DBSCAN dapat digunakan untuk mengelompokkan distribusi listrik dengan hasil yang efektif.

Kata Kunci: Analisis Data, DBSCAN, Distribusi Listrik, Pengelompokan

## 1. PENDAHULUAN

Ketersediaan listrik merupakan aspek vital dalam kehidupan manusia karena sebagian besar peralatan rumah tangga serta berbagai perangkat pendukung kehidupan mengandalkan pasokan listrik. Permintaan listrik di setiap wilayah di Indonesia berbeda-beda, dipengaruhi oleh berbagai faktor dan karakteristik unik dari masing-masing daerah. Penggunaan listrik merupakan elemen vital dalam kehidupan modern, memainkan peran krusial dalam menyokong berbagai aktivitas sehari-hari dan industri. Peran penting ini tercermin dalam konsumsi listrik yang meningkat secara signifikan di berbagai daerah di Indonesia, dipengaruhi oleh beragam faktor dan karakteristik lokal [1]. Peningkatan jumlah pelanggan dari tahun ke tahun dan kompleksitas yang semakin meningkat dalam pertumbuhan permintaan listrik di suatu daerah membutuhkan penanganan yang efisien untuk memastikan distribusi yang optimal [2]. Sebagai penyedia utama tenaga listrik untuk kepentingan umum, Perusahaan Listrik Negara (PLN) memainkan peran penting dalam memastikan akses listrik yang luas dan efisien di seluruh wilayah Indonesia [3][4].

Perseroan Terbatas Perusahaan Listrik Negara (PT. PLN) merupakan perusahaan milik negara yang berperan dalam menyediakan listrik untuk kebutuhan masyarakat di Indonesia [5][6]. PLN telah berperan penting dalam pengelolaan dan penyediaan listrik untuk berbagai sektor ekonomi dan kebutuhan masyarakat Indonesia sejak didirikan pada 27 Agustus 1965 [7]. Terlepas dari fakta bahwa pemerintah daerah bertanggung jawab untuk menjalankan fungsi regulator dalam bidang ketenagalistrikan, PLN terus diberikan prioritas utama dalam menyediakan listrik bagi masyarakat [8][9]. Perusahaan Listrik Negara (PLN) sebagai penyedia layanan distribusi listrik di Indonesia mencatat bahwa pada tahun 2020, panjang jaringan distribusi listrik Indonesia mencapai 1.006.265,15 kilometer, mengalami peningkatan sekitar 2.67% hingga 5.25% sejak tahun 2015 hingga 2020 [10][11]. Angka tersebut digunakan untuk mengukur rasio elektrifikasi di seluruh wilayah Indonesia karena memberikan gambaran tentang seberapa jauh distribusi listrik telah menjangkau wilayah-wilayah tersebut oleh PLN [12]. Rasio ini tidak hanya mencerminkan tingkat aksesibilitas terhadap listrik, tetapi juga menjadi indikator efisiensi sistem distribusi di suatu daerah [13].

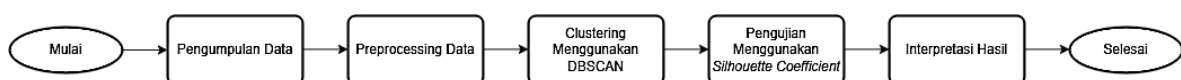
Menurut data statistik PT. PLN (Persero) tahun 2015, rasio elektrifikasi di seluruh Indonesia mencapai 78,06%. Pada akhir Desember 2015, kapasitas terpasang total dan jumlah unit pembangkit PLN adalah 40.265 MW dan 5.218 unit, dengan 25.182 MW (62,54%) berlokasi di Jawa. Kapasitas terpasang ini meningkat 0,06% dibandingkan dengan akhir Desember 2014 [1][15]. Data pelanggan yang dimiliki oleh PLN bisa digunakan untuk mengenali pola pengelompokan distribusi listrik di setiap wilayah dengan memanfaatkan teknik data mining, seperti metode clustering. Proses pengelompokan data ke dalam beberapa kelompok adalah inti dari clustering [14]. Salah satu algoritma clustering yang dapat menganalisis distribusi listrik adalah *Density Based Spatial Clustering of Applications with Noise* (DBSCAN) [16][17].

DBSCAN menawarkan kemampuan untuk mengidentifikasi *cluster* berdasarkan kerapatan spasial data, memungkinkan penyedia layanan untuk memahami pola distribusi energi dengan lebih mendalam [18][19][20]. Selain itu, DBSCAN mampu menangani anomali atau titik-titik noise yang mungkin ada dalam distribusi listrik, sehingga memberikan fleksibilitas dalam menangani struktur cluster yang kompleks [21][22]. Penelitian sebelumnya telah menunjukkan bahwa penggunaan DBSCAN dapat menghasilkan hasil yang lebih akurat dalam mengelompokkan pola distribusi listrik dibandingkan dengan metode lain [23][25].

Penelitian oleh Mustika Putri dan rekan-rekan menunjukkan bahwa metode DBSCAN lebih unggul dibandingkan K-Means dalam melakukan pengelompokan data [1][14]. Hal ini ditunjukkan oleh nilai koefisien silhouette dari algoritma DBSCAN yang lebih mendekati 1 daripada algoritma K-Means. Penelitian lainnya oleh Kaige Wang dan tim menemukan bahwa meskipun K-Means lebih cepat dan memberikan hasil yang lebih umum, DBSCAN mampu mengekstraksi lebih banyak detail dari kumpulan data [15]. Berdasarkan temuan-temuan ini, dapat disimpulkan bahwa penggunaan algoritma DBSCAN dalam analisis distribusi listrik menawarkan keunggulan signifikan dalam hal kedetailan dan kualitas pengelompokan, memberikan wawasan yang lebih mendalam mengenai distribusi listrik dan penerapan algoritma DBSCAN [16]. Penelitian ini akan akan diolah berdasarkan wilayah geografis yang mengacu pada kepadatan penduduk, geografis, dan infrastruktur. Dengan demikian, penelitian ini bertujuan untuk menggunakan algoritma DBSCAN dalam pengelompokan data distribusi listrik berdasarkan letak geografis wilayah di Indonesia.

## 2. METODOLOGI PENELITIAN

Penelitian ini mengikuti tahapan utama sebagaimana tergambar dalam Gambar 1, yang akan menganalisis pengelompokan atau clustering dengan menggunakan metode DBSCAN.



Gambar 1. Metode Penelitian

### 2.1. Tahap Pengumpulan Data

Pada proses pengumpulan data, seluruh data dan informasi yang dibutuhkan untuk penelitian dikumpulkan. Metode yang diterapkan untuk proses pengumpulan data adalah sebagai berikut:

1. Literatur: Mencari referensi terkait faktor-faktor yang mempengaruhi, termasuk penelitian sejenis dengan studi kasus yang berbeda.
2. Observasi: Melakukan observasi terhadap *website* PT. PLN (Persero) untuk mengidentifikasi faktor-faktor penting terkait data listrik di setiap wilayah. Data tersebut diolah menggunakan DBSCAN untuk mengelompokkan pendistribusian listrik.
3. Studi Pustaka: Memahami dan mencari referensi tambahan tentang penambahan pendistribusian listrik di setiap wilayah, termasuk data distribusi listrik dari [website https://web.pln.co.id/](https://web.pln.co.id/).

### 2.2. Tahap Preprocessing Data

Tahap *Preprocessing* data merupakan bagian penting dalam analisis data. Langkah ini bertujuan untuk mempersiapkan data mentah agar siap digunakan dalam proses data *mining* atau analisis lebih lanjut. Pentingnya tahap ini terletak pada peningkatan akurasi dan efisiensi model analisis melalui kualitas data yang baik. *Preprocessing* dilakukan untuk memastikan bahwa data yang digunakan dalam analisis adalah data yang valid. Pada penelitian ini menggunakan teknik *transformasi* data agar data dapat diolah sesuai dengan tujuan penelitian.

Transformasi data adalah proses mengubah data dari format atau struktur yang satu ke format atau struktur lain untuk memudahkan analisis dan pengolahan lebih lanjut. Pada tahap ini, data mentah yang dipilih diubah menjadi data yang diinginkan yaitu penjumlahan pada seluruh variabel pada tahun 2014 hingga tahun 2022. *Transformasi* ini bertujuan agar data tersebut cocok dan dapat digunakan dalam proses data mining. Melalui tahap *preprocessing*, data disiapkan dengan baik untuk analisis, memastikan bahwa hasil yang diperoleh lebih akurat dan efisien

### 2.3. Tahap Clustering Menggunakan DBSCAN

*Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) merupakan salah satu dari algoritma *clustering* yang menggunakan kepadatan dari data untuk membentuk kluster [13]. DBSCAN ini memungkinkan identifikasi kluster dengan bentuk dan ukuran yang beragam, serta dapat menangani keberadaan *noise* dalam data. Algoritma DBSCAN digunakan dalam tahap *clustering* dengan menentukan Epsilon dan MinPts. Konsep dasar DBSCAN adalah ketercapaian kepadatan dan keterhubungan kepadatan, yang bergantung pada parameter radius maksimal (epsilon) dan jumlah minimum objek dalam sebuah cluster (MinPts) [1].

1. *Epsilon (eps)*: Jarak yang menentukan kedekatan antara objek dengan objek di sekitarnya. Objek yang memenuhi nilai ini dianggap memiliki kedekatan dengan objek yang diamati.
2. *Minimum Number of Points (MinPts)*: Jumlah minimum objek yang diperlukan untuk membentuk sebuah cluster jika saling berdekatan.
3. Objek *Border*: Objek yang berdekatan dengan objek core, sedangkan objek core adalah objek yang berdekatan dengan nilai epsilon yang diberikan. Objek yang tidak termasuk core atau border disebut sebagai anomali atau pencilan.

Berikut tahapan dalam menggunakan algoritma DBSCAN.

- a. Menentukan nilai parameter *minPts* serta *epsilon*.
- b. Menentukan secara acak titik awal atau p
- c. Ulangi Langkah 3 sampai ke 5 sampai semua titik selesai di proses
- d. Hitung *epsilon* yang ketercapaian kepadatan terhadap p menggunakan

$$E(x, y) = \sqrt{\sum_{i=0}^n (X_i - Y_i)} \quad (1)$$

- e. Jika titik memenuhi *epsilon* lebih dari *minPts* maka titik p adalah *core point*
- f. Jika p merupakan titik batas serta tidak ada titik lain yang dapat dijangkau berdasarkan kepadatan dari p, kemudian proses dilanjutkan ke titik selanjutnya.

### 2.4. Pengujian Menggunakan Silhouette Coefficient

Pengujian ini bertujuan untuk menilai kualitas dan keunggulan dari pengelompokan objek ke dalam cluster. Pendekatan ini merupakan gabungan antara metode *Cohesion* dan *Separation*, yang bertujuan untuk mengevaluasi seberapa baik objek-objek dikelompokkan bersama dalam satu cluster serta seberapa jauh jarak antara cluster-cluster yang berbeda. Silhouette score dihitung untuk setiap titik data dengan membandingkan

jarak rata-rata intra-cluster (a) dengan jarak antar cluster (b) [27]. Pengujian menggunakan Silhouette Score dapat menggunakan rumus

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

Nilai silhouette score berkisar antara -1 hingga 1. Nilai positif menunjukkan bahwa titik data tersebut lebih cocok dengan clusternya sendiri, nilai 0 menunjukkan bahwa titik data tersebut berada di batas antara dua cluster, dan nilai negatif menunjukkan bahwa titik data tersebut lebih cocok dengan cluster lain.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Pengumpulan Data

Pada penelitian ini menggunakan data pendistribusian listrik PLN dari tahun 2014 hingga 2022 yang berasal dari laporan statistik PLN yang tersedia di situs web resmi <https://web.pln.co.id/>. Pada data yang disajikan, terdapat tujuh kolom data, yaitu (1) satuan PLN/provinsi, (2) jalan umum, (3) industri, (4) bisnis, (5) Gedung kantor pemerintah, (6) sosial, dan (7) rumah tangga, yang akan di kelola dengan menggunakan algoritma DBSCAN. Tabel 1 menunjukkan sebgaaian data Propinsi dan distribusi penggunaan listrik PLN.

**Tabel 1.** Data wilayah pendistribusian listrik tahun 2022

No	Satuan PLN/ Provinsi	Rumah Tangga	Industri	Bisnis	Sosial	Gedung Kantor Pemerintah	Jalan Umum
1	Wilayah Aceh	1,895.07	228.77	558.37	228.92	109.78	133.11
2	Wilayah Sumatera Utara	6,025.69	3,152.63	1,808.45	501.49	152.05	419.50
3	Wilayah Sumatera Barat	1,791.64	827.73	639.01	192.18	82.58	97.30
4	Wilayah Riau	3,007.32	2,960.77	1,236.21	260.79	115.41	111.14
5	Wilayah Kepulauan Riau	499.83	39.76	301.90	40.88	41.20	22.98
6	Wilayah Sumatera Selatan	3,295.08	1,067.48	1,017.77	234.75	116.00	147.23
....	.....	.....	.....	.....	.....	.....	.....
33	Wilayah Jawa Barat	20,871.52	25,419.11	7,610.34	1,422.88	503.81	398.45
34	Wilayah Banten	6,049.89	16,605.16	3,407.59	395.19	163.58	84.26
35	Dist. Jakarta Raya	14,824.00	4,140.34	12,539.23	1,482.92	1,396.61	195.20

#### 3.2. Preprocessing Data

Tahap *pre-processing* dilakukan sebelum analisis untuk memastikan data yang digunakan akurat. Data yang didapatkan pada penelitian ini telah akurat dan sangat minim nilai null. Data yang di peroleh pertahun tersebut kemudian disatukan pada satu tabel berdasarkan variabel seperti yang ditunjukkan pada Tabel 2.

**Tabel 2.** Akumulasi data pendistribusian listrik dari tahun 2014-2022

No	Satuan PLN/ Provinsi	Rumah Tangga	Industri	Bisnis	Sosial	Gedung Kantor Pemerintah	Jalan Umum
1	Wilayah Aceh	14,718.97	1302.87	3866.09	1565.62	855.05	1050.67
2	Wilayah Sumatera Utara	46,551.45	23630.17	13906.16	3375.31	1150.99	3698.17
3	Wilayah Sumatera Barat	14,938.30	8056.82	4514.05	1311.53	621.73	839.32
4	Wilayah Riau	22,881.68	6931.58	9077.21	1690.11	881.13	1228.73
5	Wilayah Kepulauan Riau	3,954.30	311.98	2151.17	274.38	304.17	193.92
6	Wilayah Sumatera Selatan	27,097.24	8066.36	7913.83	1617.35	916.32	1177.85
....	.....	.....	.....	.....	.....	.....	.....
33	Wilayah Jawa Barat	166,559.08	205651.98	50909.48	9347.15	3722	3105.9
34	Wilayah Banten	40,207.45	111446.98	21891.22	2311.47	1161.61	673.2
35	Dist. Jakarta Raya	125,995.80	51211.27	108466.03	11719.59	11799.27	2684.04

Pada tahap ini juga dilakukan pembagian regionalisasi data pendistribusian listrik ke dalam beberapa regional yaitu Sumatera, Jawa-Bali, Kalimantan-Sulawesi, dan Papua. Regional ini dibagi berdasarkan letak geografis wilayah dikarenakan setiap wilayah geografis memiliki karakteristik yang berberda terhadap kepadatan penduduk, infrastruktur, dan sumber daya. Dengan membagi data distribusi listrik berdasarkan wilayah geografis memungkinkan analisis yang lebih mendalam dan pengelolaan yang lebih efektif, yang pada akhirnya mendukung tujuan utama penelitian dan perencanaan strategis. Pada tabel 3-6 adalah data distribusi PLN yang telah terbagi menjadi beberapa regional.

**Tabel 3.** Wilayah Regional Sumatra.

No	Satuan PLN/Provinsi	Label
1	Wilayah Aceh	Aceh
2	Wilayah Sumatera Utara	Sumut
3	Wilayah Sumatera Barat	Sumbar
4	Wilayah Riau	Riau
5	PT PLN Batam	Batam
6	Wilayah Kepulauan Riau	Kepri
7	Wilayah Jambi	Jambi
8	Wilayah Bengkulu	Bangkulu
9	Wilayah Sumatera Selatan	Sumel
10	Wilayah Bangka Belitung	Babel
11	Distribusi Lampung	Lampung

**Tabel 4.** Wilayah Regional Jawa-Bali

No	Satuan PLN/Provinsi	Label
1	Wilayah Banten	Banten
2	Wilayah Jawa Barat	Jabar
3	Dist. Jakarta Raya dan Tangerang	Jakarta
4	Wilayah Jawa Tengah	Jateng
5	Wilayah D.I. Yogyakarta	Jogja
6	Dist. Jawa Timur	Jatim
7	Distribusi Bali	Bali
8	Wilayah Nusa Tenggara Barat	NTB
9	Wilayah Nusa Tenggara Timur	NTT

**Tabel 5.** Wilayah Regional Kalimantan-Sulawesi.

No	Satuan PLN/Provinsi	Label
1	Wilayah Kalimantan. Barat	Kalbar
2	Wilayah Kalimantan Tengah	Kalteng
3	Wilayah Kalimantan Selatan	Kalsel
4	Wilayah Kalimantan Timur dan Utara	Kaltim/Kal tara
5	PT PLN Tarakan	Tarakan
6	Wilayah Sulawesi Barat	Sulbar
7	Wilayah Sulawesi Selatan	Sulsel
8	Wilayah Sulawesi Tengah	Sulteng
9	Wilayah Sulawesi Tenggara	Sultra
10	Wilayah Gorontalo	Gorontalo
11	Wilayah Sulawesi Utara	Sulut
12	Wilayah Maluku	Maluku
13	Wilayah Maluku Utara	Malut

**Tabel 6.** Wilayah Regional Papua.

No	Satuan PLN/Provinsi	Label
1	Wilayah Papua Barat	Pabar
2	Wilayah Papua	Papua

Dalam Tabel 6, regional Papua hanya mencakup dua wilayah, yaitu Wilayah Papua Barat dan Wilayah Papua. Oleh karena itu, proses *clustering* tidak dilakukan untuk regional Papua karena jumlah wilayah yang terbatas dan tidak memungkinkan membentuk cluster. Algoritma *clustering* dirancang untuk mengidentifikasi pola dan mengelompokkan data yang serupa dalam dataset yang besar. Karena regional Papua hanya memiliki dua wilayah, tidak ada cukup data untuk membentuk beberapa *cluster* yang berbeda. *Clustering* tidak akan memberikan wawasan yang signifikan atau pemisahan dalam dataset yang kecil, karena *cluster* yang bermakna biasanya memerlukan jumlah data yang lebih besar untuk secara efektif membedakan antara kelompok yang berbeda.

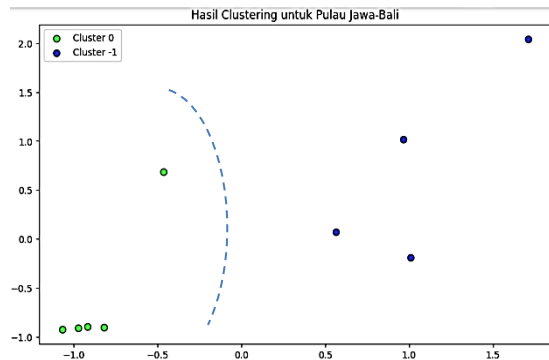
### 3.3 Algoritma DBSCAN

#### 3.3.1 Menentukan Nilai Parameter *Epsilon* dan *Min sample*

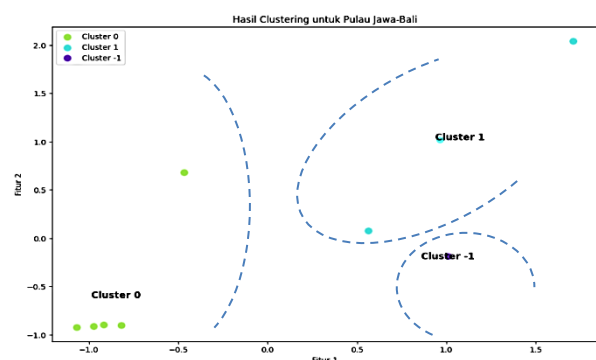
Tahap awal dalam penerapan algoritma DBSCAN adalah menentukan nilai parameter *Epsilon* dan *Min\_sample*. Pada Penelitian ini menggunakan fungsi *GridSearchCV* dari pustaka *sklearn* dari *Phyton*. Fungsi tersebut digunakan untuk menemukan parameter terbaik dari model clustering yang akan memberikan hasil clustering yang optimal [26]. Fungsi ini secara sistematis menguji berbagai kombinasi parameter yang ditentukan dalam grid dan mengevaluasi model menggunakan skor tertentu untuk memilih kombinasi parameter terbaik. Untuk clustering, "*GridSearchCV*" dapat menggunakan metrik evaluasi seperti silhouette score yang mengukur seberapa mirip objek dalam cluster dibandingkan dengan objek di cluster lain. Silhouette score memberikan nilai antara -1 dan 1, di mana nilai yang lebih tinggi menunjukkan clustering yang lebih baik.

#### 3.3.2 Clustering Menggunakan Algoritma DBSCAN

Fungsi *GridSearchCV* digunakan untuk menemukan parameter *epsilon* dan *min\_sample* terbaik berdasarkan *Shilluete score* tertinggi. Untuk ketiga region yang dievaluasi, masing-masing region terbagi ke dalam 2 cluster. Sebagai pembanding, dicoba pula tampilkan hasil dengan cluster terbanyak, yaitu masing-masing region terbagi ke dalam 3 cluster.



**Gambar 2.** Hasil Clustering regional Jawa-Bali 2 Cluster



**Gambar 3.** Hasil Clustering regional Jawa Bali 3 Cluster

**Tabel 7.** Jumlah cluster 2 regional Jawa-Bali

Cluster -1	Cluster 0
Jateng	NTT
Jatim	NTB
Jakarta	Jogja
Jabar	Bali
	Banten

**Tabel 8.** Jumlah cluster 3 regional Jawa-Bali

Cluster -1	Cluster 0	Cluster 1
Jakarta	NTT	Jateng
	NTB	Jatim
	Jogja	Jabar
	Bali	
	Banten	

Dari berbagai cluster hasil nilai epsilon dan min\_sample berbeda, anggota cluster dan jumlah cluster yang terbentuk sama. Untuk visualisasi digunakan *Principal Component Analysis* (PCA) yang memproyeksikan data dengan dimensi banyak ke dalam 2 dimensi kartesian,  $x$  dan  $y$ .

Gambar 2 dan 3 menampilkan hasil *clustering* dari regional Jawa-Bali, dimana hasil dari 2 cluster berdasarkan nilai *Silhouette Score* terbaik, yaitu 0.62. Parameter-parameter optimalnya adalah *eps* 1.7 dan *min\_pts* 4. Sedangkan Gambar 3 menggunakan parameter *eps* 1.9 dan *min\_pts* 3, dengan nilai *Silhouette Score* 0,61. Dari Gambar 2 dan 3, terlihat bahwa anggota cluster sama, kecuali Jakarta memisahkan diri dari cluster Jawa Barat, Jawa Tengah dan Jawa Timur.

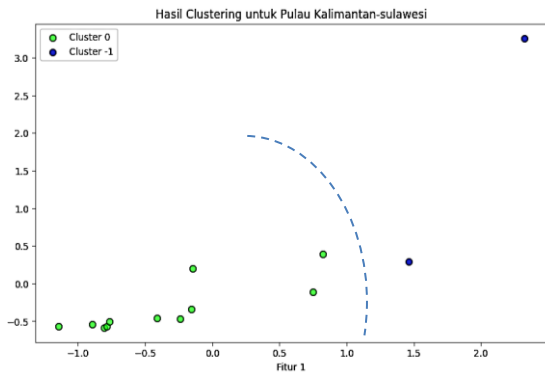
Dari analisa profil propinsinya, pembagian ke dalam 3 cluster untuk region Jawa-Bali lebih dapat menggambarkan situasi geografis, demografis, dan karakteristik daerah di sektor bisnis dan industri. Misalnya antara Jakarta dan Jawa Barat, Tengah dan Timur, memiliki kemiripan dari segi jumlah penduduk, namun sangat berbeda bila ditinjau dari kepadatan penduduk per km-nya. Di samping itu, rentang nilai dari jumlah penduduk pun sebenarnya masih ada gap yang cukup jauh antara Jakarta dan Jawa Barat/Tengah/Timur.

Sedangkan provinsi lainnya relatif sama, yaitu propinsi dengan jumlah penduduk yang jauh lebih kecil. Dari segi industri dan bisnis, profil Jakarta pun relatif berbeda dari porsi penggunaan listrik industri, karena jumlah industri di Jakarta/Tangerang tidak sebanyak di Jawa Barat/Tengah/Timur. Di Jakarta lebih banyak gedung-gedung bertingkat dibanding pabrik-pabrik yang lebih banyak mengonsumsi listrik untuk mesin-mesinnya.

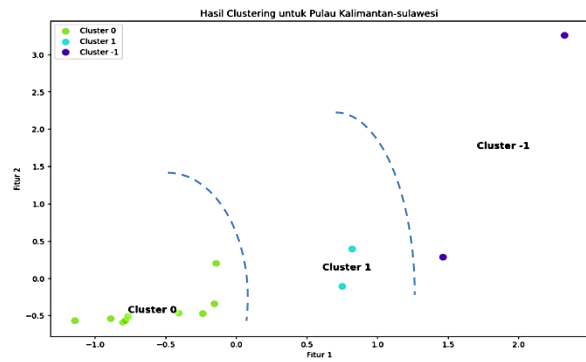
Sedangkan pada propinsi di cluster lain, secara nyata jumlahnya tidak sebanyak wilayah lainnya. Kecuali di Banten, walaupun terdapat industri pengolahan besi di Cilegon, namun tidak dapat menyaingi banyaknya pabrik dan industri yang ada di Jawa Barat/Tengah/Timur, sehingga terpisah clusternya.

Gambar 4 dan 5 juga menunjukkan perbandingan dari anggota hasil clustering yang menghasilkan 2 dan 3 cluster. Kaltim dan Kaltara, dan Sulawesi Tenggara (ibukota Makassar) memang secara nyata memiliki kota-kota industri yang besar dibandingkan wilayah lainnya. Seperti kita ketahui, di Kaltim terdapat banyak tambang minyak dan mineral, sehingga industri dan kota-kota berkembang lebih pesat dari wilayah lainnya. Demikian juga dengan Sultra, sudah sejak lama lebih maju dibanding wilayah lainnya di Sulawesi. Setelah proses clustering menghasilkan 3 cluster pun, posisi cluster yang berisi kedua propinsi tersebut tidak berubah. Sedangkan propinsi Kalbar dan Kalsel justru keluar dari cluster sebelumnya dan membentuk cluster sendiri (cluster 1). Hal ini dapat diprediksi karena kedua propinsi ini memiliki kemiripan yang lebih signifikan dibandingkan cluster pendahulunya saat masih membentuk cluster yang sama dengan cluster 0.

Gambar 6 dan 7 menunjukkan ilustrasi hasil pengelompokan untuk 2 dan 3 cluster. Terlihat bahwa propinsi Sumut konsisten berada dalam cluster terpisah. Hal ini dapat dimaklumi bahwa dari area Sumatera, Sumut sudah sejak lama berkembang industrinya, serta jumlah penduduk yang relatif lebih besar dari propinsi lainnya. Bila ditilik dari perubahan 2 ke 3 cluster, propinsi Aceh, Sumbar, Riau, Lampung dan Sumsel memisahkan diri dari cluster awalnya. Ditinjau dari data profil propinsi, anggota cluster ini, memang cenderung lebih mirip satu dengan lainnya. Namun, menurut ilustrasi Gambar 7, bisa jadi 1 atau 2 propinsi yang di sebelah kiri, dapat memiliki beberapa ciri yang sama atau mirip dengan anggota cluster 0.



**Gambar 4.** Regional Kalimantan-Sulawesi (2 Cluster)



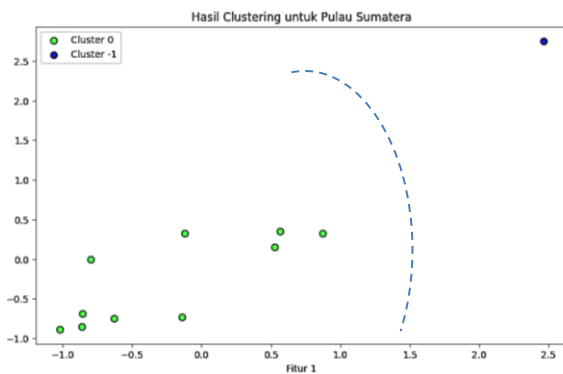
**Gambar 5.** Regional Kalimantan-Sulawesi (3 Cluster)

**Tabel 9.** Jumlah cluster 2 regional Kalimantan-Sulawesi

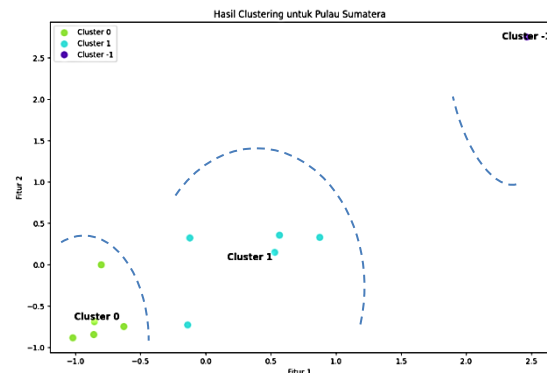
Cluster 0	Cluster -1
Sulbar	Kaltim/Kaltara
Malut	Sulsel
Maluku	
Gorontalo	
Sultra	
Sulteng	
Sulut	
Kalteng	
Kalbar	
Kalsel	
Tarakan	

**Tabel 10.** Jumlah cluster 3 regional Kalimantan-Sulawesi

Cluster 0	Cluster 1	Cluster -1
Tarakan	Kalbar	Kaltim/Kaltara
Sulbar	Kalsel	Sulsel
Malut		
Maluku		
Gorontalo		
Sultra		
Sulteng		
Kalteng		
Sulut		



**Gambar 6.** Regional Sumatera (2 Cluster)



**Gambar 7.** Regional Sumatera (3 Cluster)

**Tabel 11.** Jumlah cluster 2 regional Sumatera

Cluster 0	Cluster -1
Kepri	Sumut
Bengkulu	
Babel	
Batam	
Jambi	
Aceh	
Sumbar	
Riau	
Lampung	
Sumsel	

**Tabel 12.** Jumlah cluster 3 regional Sumatera

Cluster 0	Cluster 1	Cluster -1
Kepri	Aceh	Sumut
Bengkulu	Sumbar	
Babel	Riau	
Batam	Lampung	
Jambi	Sumsel	

**3.3. Silhouette Coefficient**

Data hasil pengujian kuantitatif dari *Silhouette Coefficient* pada algoritma DBSCAN, dapat dilihat pada Tabel 13. Tabel tersebut menunjukkan komposisi parameter DBSCAN yang menghasilkan jumlah cluster

optimal untuk setiap region. Untuk perbandingan, diberikan juga data parameter dengan hasil jumlah cluster terbanyak.

**Tabel 13.** Komposisi parameter tuning dan skor *Silhouette Coefficient*

Parameter eps	Min sample	Regional	Skor Silhouette Score	Jumlah Cluster
1.7	4	Jawa-Bali	0,62	2
1.9	3		0,61	3
1.4	3	Kalimantan-Sulawesi	0,67	2
1.5	2		0,54	3
1.9	3	Sumatera	0,64	2
1.6	3		0,57	3

Dari data pengukuran menggunakan metrik *silhouette coefficient* pada berbagai nilai parameter *epsilon* (*eps*) dan *min\_sample*, dapat disimpulkan bahwa variasi nilai *eps* dan *min\_sample* memengaruhi kualitas dan struktur kluster dalam data secara signifikan. Pada Tabel 10, untuk region Jawa-Bali mendapatkan score terbaik di angkat 0,62 dengan menggunakan parameter *eps* 1.7 dan *min\_sample* 4 dan mendapatkan cluster terbanyak dengan score 0,61 menggunakan parameter *eps* 1.9 dan *min\_pts* 3. Regional Kalimantan-Sulawesi mendapatkan score terbaik yaitu 0,67 dan juga cluster terbanyak yaitu 3 cluster dengan score 0,54. Pada regional Sumatera mendapatkan score terbaik yaitu 0,64 dan *cluster* terbanyak yaitu 3 *cluster* dengan score 0,57.

### 3.4. Perbandingan dengan Metode Lainnya

Dari data yang sama, Mulyadi dkk [28] menggunakan metode mini batch K-means clustering, dan Utari dkk [29] menerapkan metode clustering mean shift. Untuk hasil pembentukan 2 cluster, metode ini memiliki hasil yang sama dengan K-means untuk regional Jawa-Bali, dan Kalimantan-Sulawesi. Sedangkan untuk regional Sumatera, tidak dilaporkan hasil clustering metode K-means. Sedangkan untuk hasil pembentukan 3 cluster, metode K-means hanya melaporkan region Sumatera, di mana hasil yang diperoleh sama persis dengan metode yang diusulkan bila dipilih jumlah cluster terbanyak (3 cluster).

Sedangkan metode mean shift dalam [28] hanya melaporkan hasil sebanyak 3 cluster untuk semua region kecuali Papua. Untuk region Sumatera, terdapat perbedaan anggota pada cluster tengah, yaitu tidak terdapat Aceh pada metode mean-shift, sedangkan metode ini dan K-means, Aceh berada pada cluster tengah. Untuk region Jawa-Bali, hasil cluster metode ini juga berbeda dengan metode K-means, yaitu pada K-means, Jakarta menjadi cluster yang sama dengan Jabar, sedangkan untuk cluster tengah, Banten bergabung dengan Jateng dan Jatim. Terakhir untuk region Kalimantan-Sulawesi, juga terdapat perbedaan hasil, di mana pada metode K-means, Sulsel sendirian berada pada cluster terpisah, sedangkan Kaltim/Kaltara bergabung ke cluster tengah. Bila dilihat dari visualisasi pada proyeksi sumbu x dan y, posisi propinsi untuk cluster tengah memang lebih tepat diisi oleh 3 propinsi, yaitu Kalbar, Kalsel dan Kaltim/Kaltara. Besar kemungkinan untuk wilayah regional ini, metode K-means lebih tepat dalam melakukan pengelompokan. Sedangkan untuk yang lainnya, masih perlu analisa lebih mendalam dari berbagai faktor yang mungkin juga perlu dilibatkan dalam pembentukan cluster.

## 4. KESIMPULAN

Berdasarkan penelitian yang melalui tahapan identifikasi masalah, studi literatur, dan pengumpulan data dari sumber-sumber yang relevan, serta pre-processing data, penelitian ini berhasil menunjukkan bahwa data distribusi listrik dapat diklasifikasikan ke dalam beberapa cluster menggunakan DBSCAN. Algoritma ini dipilih karena kemampuannya dalam mengidentifikasi pola dan mengelompokkan data yang serupa dalam dataset yang besar, serta kemampuannya menangani anomali atau noise dalam data. Hal ini mengindikasikan bahwa terdapat perbedaan dalam pola distribusi listrik berdasarkan kategori penggunaannya di berbagai provinsi di Indonesia. Hasil analisis menunjukkan bahwa algoritma DBSCAN mampu mengelompokkan data distribusi listrik di Indonesia dengan baik. Algoritma DBSCAN mampu mengelompokkan data distribusi listrik di Indonesia menjadi 2 hingga 3 kluster berdasarkan letak geografis. Nilai Silhouette Score yang diperoleh menunjukkan bahwa kualitas dan struktur kluster dalam data secara signifikan. Nilai Silhouette Score terbaik untuk regional Jawa-Bali adalah 0,62 dengan menggunakan parameter *eps* 1.7 dan *min\_sample* 4, sedangkan untuk regional Kalimantan-Sulawesi adalah 0,67 dengan parameter *eps* 1.4 dan *min\_sample* 3, dan untuk regional Sumatera adalah 0,64 dengan parameter *eps* 1.9 dan *min\_sample* 3. Hasil clustering menunjukkan bahwa terdapat perbedaan pola distribusi listrik di antara regional geografis di Indonesia. Secara umum, penelitian ini menunjukkan bahwa DBSCAN efektif sebagai alat untuk menganalisis distribusi listrik di Indonesia. Ini memberikan kontribusi dalam membuat keputusan yang lebih baik untuk perencanaan dan pengelolaan distribusi energi dan memahami pola distribusi listrik di Indonesia berdasarkan letak geografis.





- [20] P. N. Lhokseumawe, K. Pengantar, rahayu deny danar dan alvi furwanti Alwie, A. B. Prasetyo, and R. Andespa, "Prediksi Beban Listrik Menggunakan Metode Fuzzy Time Series Dan Particle Swarm Optimization (Studi Kasus: Pt Pln Rayon Panam)," *J. Ekon. Vol. 18, Nomor 1 Maret201*, vol. 2, no. 1, pp. 41–49, 2020.
- [21] N. R. Hikmiah, R. R. A. Siregar, B. Prayitno, D. T. Kusuma, and N. G. Pahiyanti, "Metode Fuzzy Subtractive Clustering Dalam Pengelompokan Penggunaan Energi Listrik Rumah Tangga," *Petir*, vol. 14, no. 2, pp. 269–279, 2021, doi: 10.33322/petir.v14i2.1448.
- [22] G. C. Pamuji and H. Rongtao, "A Comparison study of DBScan and K- Means Clustering in Jakarta rainfall based on the Tropical Rainfall Measuring Mission (TRMM) 1998-2007," *IOP Conf. Ser. Mater. ...*, 2020, doi: 10.1088/1757-899X/879/1/012057.
- [23] S. Chowdhury, N. Helian, and R. C. de Amorim, "Feature weighting in DBSCAN using reverse nearest neighbours," *Pattern Recognition*. Elsevier,2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0031320323000158>
- [24] Y. Xu, Y. Zheng, C. He, and Z. Wang, *2019 Chinese Control And Decision Conference (CCDC)*, no. 2019 Ccdc. 2019.
- [25] P. PLN (PERSERO), "Company Profile Company Profile," *Mki*, no. 16, pp.1–17, 2019, [Online]. Available: <https://www.usf.edu/business/documents/departments/finance/smif/analysis-baba.pdf%0Ahttps://stories.starbucks.com/uploads/2019/01/AboutUs-Company-Profile-1.6.21-FINAL.pdf>
- [26] AHMAD, Ghulab Nabi, et al. Efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV. *IEEE Access*, 2022, 10: 80151-80173.
- [27] Li, Z., Zhang, Y., & Wu, X. (2020). Silhouette-based clustering for time series data. In 2020 IEEE International Conference on Data Science (ICDS) (pp. 274-279). IEEE.
- [28] S. Mulyadi, F. Insani, S. Agustian, and L. Afriyanti, "Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Mini Batch K-Means," vol. 3, 2024.
- [29] R. Utari, F. Insani, S. Agustian, and L. Afriyanti, "Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Mean shift," vol. 3, 2024.