

KLASTERISASI PADA DATA PENGGUNAAN LISTRIK DI GEDUNG TELKOM UNIVERSITY MENGGUNAKAN ALGORITMA DENSITY- BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE (DBSCAN)

CLUSTERING ON ELECTRICITY USAGE AT TELKOM UNIVERSITY BUILDING USING DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE (DBSCAN)

Reza Mahendra¹, Fairuz Azmi², Casi Setianingsih³

^{1,2,3} Universitas Telkom, Bandung

rzmhndra@student.telkomuniversity.ac.id¹, worldliner@telkomuniversity.ac.id²,

casisetianingsih@telkomuniversity.ac.id³

Abstrak

Proses pengubahan energi listrik tentunya membutuhkan biaya yang sesuai dengan jumlah pemakaiannya terutama pada tempat-tempat yang membutuhkan energi listrik dalam jumlah besar. Dengan jumlah konsumsi listrik yang besar, biaya yang dikeluarkan juga akan besar. Namun karena sulitnya memonitoring pemakaian listrik secara manual, ditambah tagihan listrik yang tidak menampilkan detail pemakaiannya, maka dengan menggunakan metode clustering dan menganalisa jumlah pemakaian listrik akan memudahkan dalam monitoring pemakaian listrik sehingga dapat dikelola dengan baik. Hasil dari penelitian ini menggunakan algoritma DBSCAN menghasilkan nilai Silhouette Coefficient untuk data pengujian per bulan menggunakan data real dan data perangkat virtual sebesar 1,0, data pengujian per hari sebesar 1,0 untuk data nyata dan 0,86 untuk data perangkat virtual dan data pengujian per gedung 0,86.

Kata kunci : Penggunaan Listrik, Monitoring, DBSCAN, Clustering

Abstract

The process of converting electrical energy certainly requires costs that are in accordance with the amount of use, especially in places that require large amounts of electrical energy. With a large amount of electricity consumption, the costs incurred will also be large. However, because of the difficulty of monitoring electricity consumption manually, plus the electricity bill does not display the details of its usage, then using the clustering method and analyzing the amount of electricity consumption will make it easier to monitor electricity consumption so that it can be managed properly. The results of this research using the DBSCAN algorithm produce a Silhouette Coefficient value for testing data per month using real data and virtual device data of 1.0, testing data per day of 1.0 for real data and 0.86 for virtual device data and testing data per building of 0.86.

Keywords: Electricity, Monitoring, Clustering, DBSCAN

1. Pendahuluan

Listrik merupakan salah satu sumber energi yang sangat penting peranannya dalam kehidupan sehari-hari bahkan sebagai kebutuhan hidup [1]. Hampir setiap pengguna tidak dapat memantau besarnya konsumsi listrik karena tidak mudah untuk mengetahui detail penggunaan listrik [2].

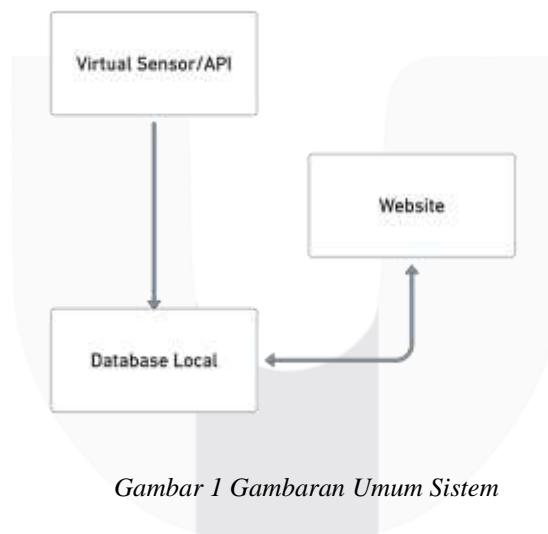
Penggunaan listrik dalam jumlah besar hampir selalu terdapat pada bangunan atau tempat-tempat yang membutuhkan energi listrik dalam jumlah besar, salah satunya adalah gedung di lingkungan kampus. Kampus ini memiliki konsumsi listrik yang cukup besar karena kegiatan di kampus tidak hanya pada jam belajar mengajar, bahkan di luar jam tersebut. Akibatnya, konsumsi listrik cukup sulit untuk dipantau dan dapat mengakibatkan tagihan yang besar.

Penelitian tentang clustering sangat sering ditemukan dengan menggunakan algoritma clustering K-means. Penelitian ini berfokus pada penggunaan algoritma Density-Based Spatial

Clusterieng Application with Noise (DBSCAN) yang diharapkan akan mendapatkan hasil cluster yang lebih baik, karena DBSCAN tidak meminta pengguna untuk memasukkan nilai berapa banyak cluster yang akan dibuat sehingga akan membuat lebih mudah terutama bagi yang bertugas memeriksa pemakaian listrik langsung ke gedung-gedung di Telkom University.

2. Perancangan Sistem

Pada perancangan sistem ini akan dibangun *web* yang dapat menampilkan informasi terkait data listrik untuk memudahkan tim logistik monitoring listrik di gedung Telkom University. Sistem ini pada dasarnya akan mengklaster data penggunaan listrik dari data mentah menjadi data yang sudah siap untuk diklaster dengan menggunakan beberapa algoritma, pada penelitian ini berfokus pada algoritme DBSCAN.



Gambar 1 Gambaran Umum Sistem

Sistem terbagi atas 3 bagian dan terhubung oleh internet, Virtual Sensor berfungsi untuk mengirimkan data penggunaan listrik ke *database*, *website* berfungsi untuk menampilkan informasi berupa data listrik seperti KWH, delta KWH ataupun kategori *cluster* dengan menggunakan data yang sudah dikirim oleh *virtual sensor* melalui *database* menggunakan algoritme DBSCAN.

2.1. Sistem Pemantauan Penggunaan Listrik

Sistem pemantauan listrik memiliki arti penting khususnya dalam mengatur manajemen penggunaan listrik dalam suatu gedung. Banyak penelitian menunjukkan bahwa ini memiliki efektivitas untuk mengurangi penggunaan listrik di suatu tempat atau gedung, khususnya dengan jumlah penggunaan yang besar [3]. Listrik setiap hari adalah kebutuhan hampir untuk semua manusia demi berjalannya aktivitas yang dilakukan setiap harinya.

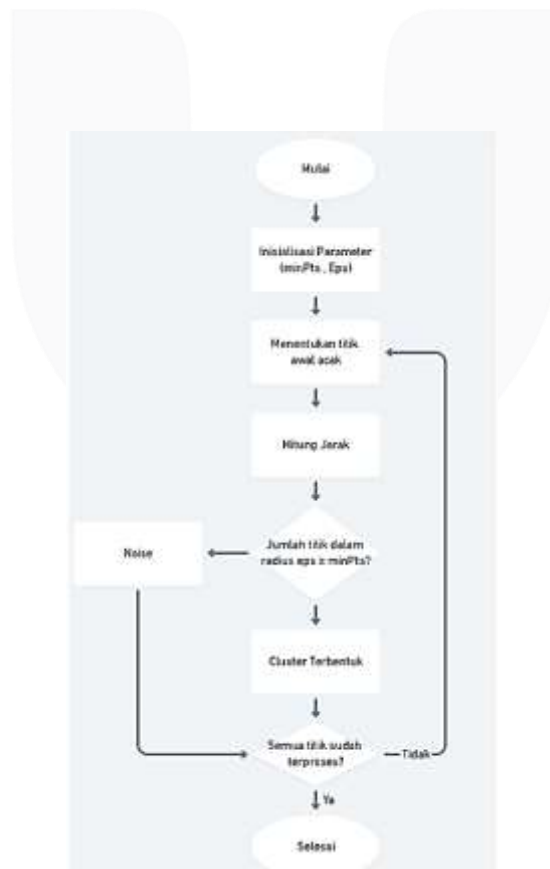
Sistem pemantauan atau *monitoring* listrik ini adalah sistem pengecekan secara jarak jauh yang memudahkan tim logistik Telkom University. Adanya penelitian ini karena listrik memegang peran penting umumnya pada gedung-gedung yang menggunakan listrik yang cukup besar seperti kampus, terlebih penggunaan listrik di Telkom University bahkan sering digunakan di luar jam belajar mengajar sekalipun. Untuk dapat manajemen penggunaan listrik yang baik dibutuhkan suatu sistem yang dapat memantau penggunaan listrik sehingga penggunanya bisa manajemen penggunaan listrik dengan lebih baik [4].

2.2. Clustering

Clustering adalah teknik data mining yang mengelompokkan objek-objek yang serupa [5]. Clustering mengelompokkan sekumpulan titik data sedemikian rupa sehingga terdapat kesamaan antara cluster tinggi dan rendah [6] atau algoritma clustering mencari kesamaan atau ketidakmiripan antar titik data. Metode yang sering digunakan untuk mengukur kemiripan data adalah dengan menghitung jarak antar data menggunakan Euclidean Distance, dengan persamaan sebagai berikut:

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (1)$$

Ada juga beberapa metode lain untuk mengukur jarak, tetapi jarak Euclidean adalah yang paling umum digunakan. Adapun algoritma clustering yang digunakan dalam penelitian ini adalah DBSCAN, berikut adalah flowchart dari DBSCAN:



Gambar 2 Diagram Alir DBSCAN

2.3. DBSCAN

DBSCAN merupakan salah satu algoritma dari Metode Pengelompokan Berbasis Kepadatan. Density Based sendiri merupakan metode yang dikembangkan berdasarkan kepadatan tertentu. Metode ini mendefinisikan cluster sebagai area yang berisi objek padat atau kepadatan tinggi, yang dipisahkan oleh area kepadatan rendah (merepresntasikan noise). [7].

Algoritma ini membutuhkan dua parameter input untuk dapat melakukan cluster yaitu Eps untuk radius yang menentukan batas lingkungan dari titik (Eps-neighborhood) dan MinPts, jumlah titik minimum yang harus ada di Eps-neighborhood. Urutan algoritma dari DBSCAN umumnya memiliki 6 langkah:

- 1) Poin p awal terpilih secara acak
- 2) Ambil semua poin yang *density reachable* terhadap titik p
- 3) Jika p adalah *core* poin maka *cluster* terbentuk
- 4) Jika p adalah *border* poin, tidak ada yang merupakan hubungan *density-reachable* dari p dan DBSCAN akan mengunjungi poin selanjutnya dari *database*
- 5) Lanjutkan proses sampai semua poin telah diproses

Hasil yang didapatkan tidak tergantung dari urutan dari proses poin yang diambil.

2.4. Density Based Method

Density-Based Methods merupakan kumpulan metode yang dikembangkan berdasarkan *density* tertentu. Metode ini menganggap cluster sebagai suatu area yang berisi objek-objek yang padat atau sesak, yang dipisahkan oleh area yang memiliki kepadatan rendah (merepresentasikan *noise*) [8]. Adapun beberapa metode dalam DBSCAN yaitu sebagai berikut:

1. Directly Density Reachable

Sebuah titik dikatakan *directly density-reachable* dari titik lainnya jika jarak di antara mereka tidak lebih dari nilai Eps. Titik q dikatakan *directly density-reachable* dari titik p jika titik q adalah \in Neps(p) dan p adalah *core* poin. Jarak dari titik ke titik lainnya tidak lebih dari nilai Eps [9].

$$\text{if: } \{q \text{ belongs to Neps}(p) : | \text{dist}(p, q) \leq \text{Eps}\}$$

dimana kondisi core point harus melebihi dari parameter minpts yang ditentukan.

2. Density Reachable

Sebuah titik item dikatakan *density-reachable* dari titik item yang lain jika ada suatu rantai yang menghubungkan keduanya yang berisi hanya titik-titik yang *directly density-reachable* dari titik-titik sebelumnya.

Suatu objek p adalah *density reachable* dari objek q dengan respek ke ϵ dan *MinPts* dalam suatu set objek D jika terdapat suatu rantai objek p_1, p_2, \dots, p_n , dimana $p_1 = q$ dan $p_n = p$, di mana p_{i+1} *density reachable* secara langsung dari p_i dengan respek ke ϵ dan *MinPts*, untuk $1 \leq i \leq n$, p_i anggota D.

3. Density Connected

Sebuah obyek p adalah *density-connected* terhadap obyek q dengan memperhatikan Eps dan *MinPts* dalam set obyek D, jika ada sebuah obyek o elemen D sehingga p dan q keduanya *density-reachable* dari o dengan memperhatikan Eps dan *MinPts*.

2.5. Silhouette Coefficient

Silhouette Coefficient adalah metrik yang digunakan untuk menghitung kebaikan suatu teknik *clustering* yang nilainya berkisar dari -1 sampai 1, ini juga termasuk salah satu pendekatan yang populer digunakan dalam *clustering* [9].

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

Rumus di atas dijelaskan dimana $a(i)$ adalah jarak rata-rata titik i ke semua titik lain pada data A, dan $b(i)$ adalah jarak rata-rata titik i ke titik-titik pada cluster terdekat A [10].

Berikut adalah tabel nilai struktur dari silhouette coefficient yang menandakan struktur pada sebuah cluster tersebut:

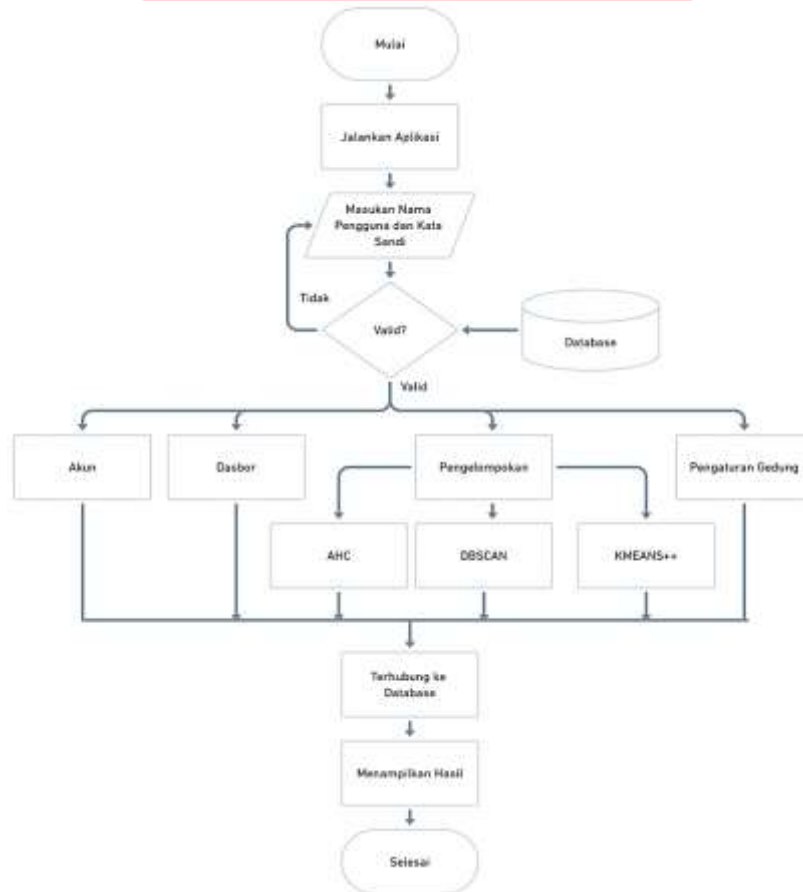
Tabel 1 Struktur Silhouette

Score	Interpretasi Silhouette Coefisien
0,71 – 1,00	Solid Structure
0,51 – 0,70	Good Structure
0,26 – 0,50	Low Structure

Score	Interpretasi Silhouette Coefisien
$\leq 0,25$	Bad Structure

2.6. Diagram Alir Sistem

Sistem *monitoring* penggunaan listrik di gedung Telkom *University* yang berbasis aplikasi *website* memiliki *flow chart* sebagai berikut:



Gambar 3 Diagram alir sistem

Sistem ini dimulai dengan melakukan login pada web client. Jika login berhasil, maka akan langsung masuk ke dashboard yang sekaligus menjadi halaman utama dan terdapat beberapa fitur yang dapat digunakan yaitu, Akun, Dashboard, Cluster dan pengaturan gedung. Pada fitur pengaturan yang khusus untuk super admin, admin dapat menghapus, menambah, mengganti password dan melihat data user, menambah kan gedung, melihat user dan gedung yang sudah terdaftar di database sistem.

Selanjutnya pada halaman dashboard, terdapat hasil cluster antar gedung yang menunjukkan hasil clustering dan disajikan dalam card-card yang memiliki warna-warna sesuai indikator kelas nya, serta terdapat juga fitur *clustering* pada *side bar*. Selanjutnya jika masuk pada fitur clustering, terdapat 3 algoritma yaitu AHC, DBSCAN dan KMEANS++ dalam menentukan tingkat kenaikan data penggunaan listrik melalui pendekatan *clustering*. Setelah memilih algoritma dan rentang bulan atau hari, hasil clustering akan muncul dan disajikan dalam bentuk pie/doughnut chart, serta menampilkan detail kWh di tiap jam atau hari nya.

3. Pembahasan

3.1. Preprocessing Data

Ada tiga algoritma yang dapat digunakan dalam sistem ini, namun dalam penelitian ini fokus pada DBSCAN. Data mentah yang diambil menggunakan API akan diproses terlebih dahulu sebelum dilakukan proses clustering. berikut adalah data sebelum dan sesudah preprocessing dan juga mendapatkan nilai delta yang nantinya akan di cluster

Tabel 2 Data Sebelum Preprocessing

No	Date	Time	Device	Kwh
0	2021-02-16	21:12:43	5	0.0
1	2021-02-16	21:13:08	5	0.0
2	2021-02-16	21:13:33	5	0.0
3	2021-02-16	21:13:57	5	0.0
4	2021-02-16	21:14:22	5	0.0
...
31898	2021-08-01	10:06:16	5	405.0

Tabel 3 Data Setelah Preprocessing

Datetime	Kwh	Old Kwh	Delta Kwh
2021-04-29 16:00:00	208.0	207.0	1.0
2021-04-30 09:00:00	209.0	208.0	1.0
2021-04-30 10:00:00	211.0	209.0	2.0
2021-04-30 11:00:00	213.0	211.0	2.0
...
2021-08-01 15:00:00	405.0	404.0	1.0

Tujuan dari proses pre-processing adalah untuk menyajikan data sebelum dimasukkan ke dalam model cluster. Hal ini agar nantinya cluster dapat berjalan dengan baik, karena terdapat beberapa hal seperti anomali data pada tool sehingga pada pre-processing ini berfungsi untuk mengatasi permasalahan tersebut.

Hasil clustering data diatas akan menghasilkan 3 kelas yaitu low, medium dan high yang akan ditampilkan tergantung dari indeks warna yang ada pada sistem.

3.2. Clustering Data

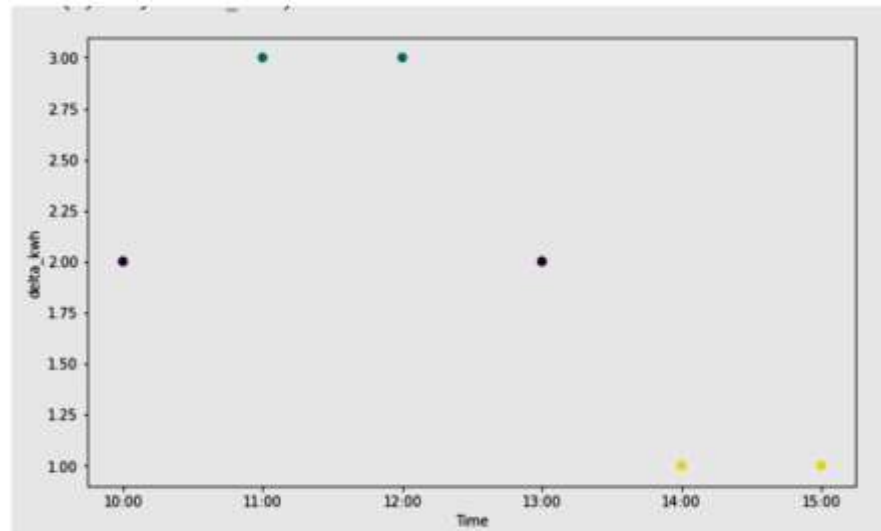
Clustering dilakukan menggunakan library scikit-learn dengan algoritma DBSCAN dengan memasukkan dua parameter yang dibutuhkan sebagai input. DBSCAN akan mencari cluster antar poin berdasarkan kepadatan melalui langkah-langkah yang sudah dijelaskan sebelumnya, sampai iterasi terakhir hingga seluruh poin sudah terproses.

Pada contoh di bawah ini terdapat data perhari pada tanggal 25 Mei 2021 pukul 10:00 – 15:00:

Tabel 4 Contoh Data

Datetime	Delta Kwh
2021-05-25 10:00:00	2.0
2021-05-25 11:00:00	3.0
2021-05-25 12:00:00	3.0
2021-05-25 13:00:00	2.0
2021-05-25 14:00:00	1.0
2021-05-25 15:00:00	1.0

Dan akan dilakukan proses clustering dengan algoritma DBSCAN, dengan ketentuan parameter sudah ditentukan. Maka akan menghasilkan grafik sebagai berikut:

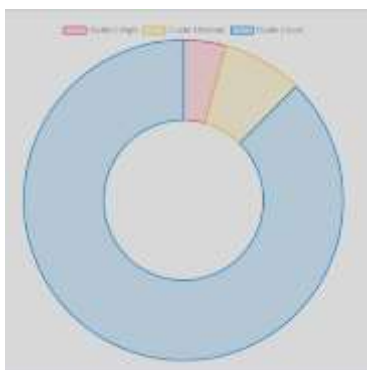


Gambar 4 Hasil Grafik Clustering

Pada gambar di atas terdapat 3 cluster, yaitu biru, ungu dan kuning. Masing-masing warna mengindikasikan bahwa kuning adalah rendah, ungu sedang dan biru tinggi. Algoritma DBSCAN tidak mengharuskan penggunaanya untuk mengatur berapa klaster yang akan terjadi nantinya, karena algoritma ini dapat mencari sendiri. Karena karakteristik dari algoritma DBSCAN seperti yang dijelaskan di atas, maka dalam beberapa kasus algoritma ini menemukan kurang dari 3 cluster atau pun lebih dari 3 cluster tergantung dari data-data yang digunakan.

3.3. Visualisasi Data

Data yang sebelumnya sudah dipreprocessing dan sudah dites melalui pandas menghasilkan grafik selanjutnya akan divisualisasi dengan menampilkan grafik pie chart dan detail penggunaan listrik pada sistem monitoring ini. Berikut adalah hasilnya:



Gambar 5 Hasil perbulan data real



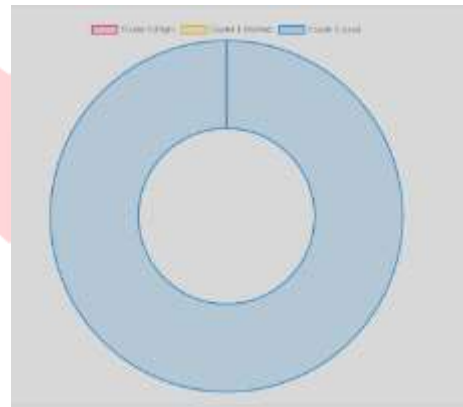
Gambar 6 Hasil perbulan virtual device

Pada gambar 5 dan 6 menunjukkan menggunakan data real dan data device periode perbulan, pada data real menggunakan bulan may, sedangkan data virtual menggunakan bulan Juli. keduanya mendapatkan hasil cluster yaitu biru rendah, kuning sedang dan merah tinggi

yang mengindikasikan tingkat penggunaan kwh. Adapun silhouette score dari hasil perbulan ini sama-sama mendapatkan skor 1.0 yang berarti termasuk struktur kuat.



Gambar 7 Hasil perhari data real



Gambar 8 Hasil perhari virtual device

Pada gambar 7 dan 8 menunjukkan menggunakan data real dan data device periode perhari, pada data real menggunakan data pada tanggal 19 Mei, sedangkan data virtual menggunakan data pada tanggal 19 Juli. Pada hasil data real didapat 3 cluster seperti sebelumnya dan mendapat nilai silhouette score sebesar 1.0 yang berarti struktur kuat. Sedangkan untuk hasil perhari data virtual, hanya mendapatkan 1 cluster karena data pada tanggal 19 Juli ini hanya sedikit, dengan silhouette score sebesar 0.86.



Gambar 9 Hasil Antar gedung

Sedangkan hasil yang diperoleh seperti di atas untuk hasil cluster pergedung dengan indikator biru kelas rendah, hijau normal dan merah tinggi. Dengan nilai silhouette coefficient sebesar 0,86 yang berarti nilai struktur clusternya kuat.

4. Kesimpulan

Berdasarkan hasil percobaan, dapat diambil beberapa kesimpulan:

1. Clustering dapat digunakan untuk mengelompokkan penggunaan data listrik
2. Pada penelitian ini, hasil skor siluet 1 dan 0,86 masih dalam struktur yang kuat.

3. DBSCAN akan menentukan kelas yang dihasilkan itu sendiri, sehingga pengguna hanya dapat mengatur dari parameter input
4. Dengan demikian, berdasarkan poin 3, dalam beberapa kasus DBSCAN bisa mendapatkan cukup 3 kelas atau bahkan lebih.
5. Hasil yang diperoleh juga akan terlihat bagus dan cocok untuk algoritma ini jika datanya lebih banyak daripada jumlah data yang sedikit.

Referensi

- [1] A. Saleh, "Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Citec Journal*, vol. 2, p. 207, 2015.
- [2] M. A. H. M. Isa, M. F. Abdul Latip, N. Zaini and Y. F. Alias, "Android-based Application for real Time Energy Monitoring of Domestic Electricity," *IEEE*, p. 134, 2015.
- [3] Chupong, C., & Plangklang, B. (2017). Electricity bill forecasting application by home energy monitoring system. 2017 International Electrical Engineering Congress (iEECON).
- [4] Aryani. Ressi "Pembangunan Purwarupa Sistem Pemantauan dan Pengelompokan Data Menggunakan *K-Means Clustering* Pada Sistem Pemetaan KWH Meter Berbasis IOT," Bandung, Telkom University, 2019.
- [5] Vijayalakshmi, S., & Punithavalli, M. (2010). *Improved varied density based spatial clustering algorithm with noise. 2010 IEEE International Conference on Computational Intelligence and Computing Research*
- [6] Nazari, Z., Kang, D., Asharif, M. R., Sung, Y., & Ogawa, S. (2015). A new hierarchical clustering algorithm. 2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS).
- [7] D. M Ester, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, 2nd International Conference on Knowledge Discovery and Data Mining, 1996.
- [8] A. O. Quintana, "Clustering: Density-based Clustering", Universidad San Pablo, p. 4, July, 2014 [Online].
- [9] M. Belal Al- Zoubi dan M. al Rawi, "An Efficient Approach for Computing Silhouette Coefficients," *J. Comput. Sci.*, vol. 4, no. 3, hal. 252–255, 2008.
- [10] P.J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53-65, 1987.

