



Grouping Electricity Distribution Data Using The Mini Batch K-Means Clustering Algorithm

Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Mini Batch K-Means Clustering

Syahrul Mulyadi¹, Fitri Insani^{2*}, Surya Agustian³, Liza Afriyanti⁴

^{1,2,3,4}Program Studi Teknik Informatika, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

E-Mail: ¹syahrulmulyadi28@gmail.com, ²fitri.insani@uin-suska.ac.id,
³sagustian@uin-suska.ac.id, ⁴liza.afriyanti@uin-suska.ac.id

Received May 4th 2024; Revised Jun 18th 2024; Accepted Jun 23th 2024
Corresponding Author: Fitri Insani

Abstract

Electricity distribution management is an important aspect of infrastructure that requires a deep understanding of distribution patterns in different regions to ensure a stable electricity supply for the community and the government. However, managing electricity distribution data involving various variables such as households, industries, businesses, social, government office buildings, and public roads requires an effective approach in grouping data to identify significant patterns. This research aims to implement the k-means mini batch algorithm to group electricity distribution in each region. The data used in this study is electricity distribution data from 2014-2022 consisting of 35 provinces. The test results showed that there were three significant clusters in the researchers' data, each marked with a different Silhouette Score. The cluster with the highest Silhouette Score is Cluster 2 with a value of 0.625, indicating high cohesion in the Kalimantan-Sulawesi regional area observed in this cluster 2. Meanwhile, Cluster 4 has the lowest Silhouette Score with a score of 0.419, indicating a lower level of cohesion in the electricity distribution pattern in each regional region of Kalimantan-Sulawesi. This study confirms that the k-means mini batch algorithm is effective for grouping electricity distribution data with satisfactory results in the separation of different clusters.

Keyword: Clustering, Data Mining, Distribution, Electricity, Mini Batch K-means

Abstrak

Manajemen distribusi listrik merupakan aspek penting dalam infrastruktur yang memerlukan pemahaman mendalam tentang pola distribusi di berbagai wilayah untuk memastikan pasokan listrik stabil bagi masyarakat dan pemerintah. Namun, mengelola data distribusi listrik yang melibatkan berbagai variabel seperti rumah tangga, industri, bisnis, sosial, gedung kantor pemerintah, dan jalan umum membutuhkan pendekatan yang efektif dalam mengelompokkan data untuk mengidentifikasi pola-pola yang signifikan. Penelitian ini bertujuan mengimplementasikan algoritma mini batch k-means untuk mengelompokkan distribusi listrik di setiap wilayah. Data yang digunakan dalam penelitian ini adalah data distribusi listrik dari tahun 2014-2022 yang terdiri dari 35 provinsi. Hasil pengujian menunjukkan bahwa terdapat tiga cluster yang signifikan dalam data peneliti, masing-masing ditandai dengan Silhouette Score yang berbeda. Cluster dengan Silhouette Score tertinggi adalah Cluster 2 dengan nilai 0.625, menunjukkan kohesi yang tinggi di wilayah regional Kalimantan-Sulawesi yang diamati dalam cluster 2 ini. Sementara itu, Cluster 4 memiliki Silhouette Score yang terendah dengan nilai 0.419, menunjukkan tingkat kohesi yang lebih rendah dalam pola distribusi listrik di setiap wilayah regional bagian Kalimantan-Sulawesi. Penelitian ini menegaskan bahwa algoritma mini batch k-means efektif untuk mengelompokkan data distribusi listrik dengan hasil yang memuaskan dalam pemisahan klaster yang berbeda.

Kata Kunci: Clustering, Data Mining, Distribusi, Listrik, Mini Batch K-means

1. PENDAHULUAN

Energi listrik sangat penting bagi kehidupan manusia karena berfungsi sebagai sumber listrik utama untuk berbagai barang yang kita andalkan, termasuk televisi, setrika, mesin cuci, telepon seluler, dan berbagai

peralatan lainnya. Energi listrik sangat meningkatkan dan memungkinkan aktivitas manusia sehari-hari, berfungsi sebagai instrumen yang sangat bermanfaat yang menyederhanakan banyak tugas [1]. Di Indonesia, terutama di perkotaan, orang-orang sudah sangat mengandalkan listrik untuk menjalani hidup mereka [2]. Seiring meningkatnya penggunaan listrik, data alokasi listrik harus dikelompokkan agar pengguna dapat memantau penggunaan listrik secara real time untuk mengoptimalkan penggunaannya [3].

Perusahaan Listrik Negara (PLN), sebuah badan usaha milik negara, memiliki peran krusial dalam menyediakan listrik bagi penduduk Indonesia [4]-[5]. Meskipun pemerintah daerah bertugas sebagai regulator di sektor ketenagalistrikan, PLN tetap mendapat prioritas utama dalam menyediakan layanan listrik kepada masyarakat [6]. Pada tahun 2020, PLN mencatat bahwa panjang jaringan distribusi listrik di Indonesia mencapai 1.006.265,15 kilometer, menunjukkan peningkatan sebesar 2.67% hingga 5.25% sejak tahun 2013 [7]-[8]. Data ini digunakan untuk menilai tingkat elektrifikasi di seluruh wilayah Indonesia, mencerminkan sejauh mana listrik telah tersedia dan sistem distribusi PLN beroperasi dengan efisien di berbagai daerah [9]. Saat ini, bagaimanapun tidak semua daerah di Indonesia memiliki akses ke listrik. Penyebab masalah ini berasal dari letak geografis suatu wilayah di provinsi tertentu, bersama dengan beberapa faktor lain yang membuat wilayah tersebut sulit untuk dijangkau. Ini terlihat dari perbedaan angka persentase elektrifikasi di setiap wilayah [10].

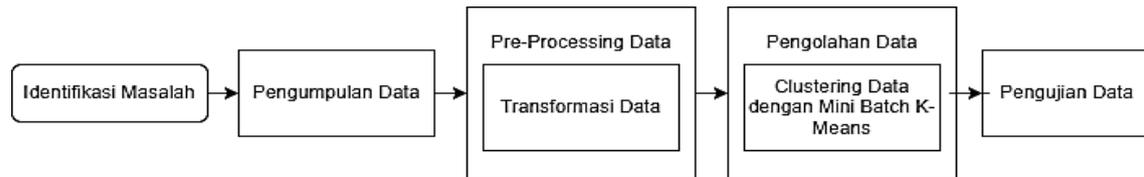
Menurut data PT PLN (Persero) pada tahun 2013, sekitar 80% wilayah di Indonesia sudah terkoneksi dengan jaringan listrik. Pada akhir Desember 2013, kapasitas total pembangkit listrik yang terpasang mencapai 34.206 MW dengan hampir 5.000 unit pembangkit, di mana sekitar 78% berada di Pulau Jawa. Ada peningkatan sekitar 4% dalam kapasitas terpasang dibandingkan dengan tahun sebelumnya [11]. Terdapat peningkatan jumlah pelanggan setiap tahun, dan semakin kompleksnya pertumbuhan ketenagalistrikan di suatu wilayah memerlukan proses yang kompleks untuk menjamin distribusi tenaga listrik yang optimal [12]. Para peneliti bermaksud menggunakan teknik ilmu komputer dan algoritma clustering untuk mengkategorikan lokasi distribusi listrik di Indonesia [13]. Beragam bidang dalam ilmu komputer yang memiliki kemampuan untuk menangani masalah-masalah yang kompleks, termasuk melakukan prediksi, estimasi, peringkat, klasifikasi, asosiasi, dan sejenisnya. Salah satu subdisiplin dalam bidang ini adalah Kecerdasan Buatan, yang mencakup bidang-bidang seperti Penambangan Data, Sistem Pendukung Keputusan, sistem pakar, Jaringan Syaraf Tiruan, dan masalah terkait lainnya [14]-[15]-[16]. Tujuan penelitian ini adalah untuk menganalisis kemungkinan penerapan teknik ilmu komputer dapat digunakan untuk mengelompokkan daerah-daerah di Indonesia, terutama dalam konteks distribusi listrik. Penelitian ini diharapkan dapat memberikan manfaat besar bagi pemerintah dalam upayanya menjamin penyediaan listrik yang merata dan mencukupi bagi seluruh masyarakat dalam bentuk jurnal akademik [17]-[18]-[19]. Oleh karena itu, teknik yang diterapkan adalah Data Mining dengan metode *Mini Batch K-means*. Algoritma Mini Batch K-Means menghemat biaya komputasi dengan tidak memproses seluruh dataset pada setiap iterasi, melainkan menggunakan sampel kecil dengan ukuran tetap. Dengan cara ini, jumlah perhitungan jarak periterasi dapat dikurangi [20]-[21]-[22].

Penelitian sebelumnya telah menunjukkan bahwa pemanfaatan algoritma Mini Batch K-Means telah meningkatkan kinerja algoritma K-Means dalam studi segmentasi gambar, khususnya pada dataset MNIST. Peningkatan ini terutama penting dalam mengurangi waktu komputasi yang diperlukan untuk mengsegmentasi gambar tulisan tangan pada kumpulan data MNIST Mini Batch K-Means, sekaligus mempertahankan tingkat akurasi yang sama. Persentasenya menurun dari 91,14% menjadi 90,13% sebesar 1,01% [23]. Selain itu, ada studi lain yang menggunakan metode K-Means untuk menyelesaikan masalah serupa, salah satunya dilakukan oleh Cici Astria [24]. Temuan penelitian menunjukkan bahwa K-Means dapat memfasilitasi penilaian distribusi listrik di berbagai wilayah di Indonesia, yang dimana menjelaskan tentang penggunaan dua kelompok centroid data dengan tingkat penyebaran yang tinggi, yang memiliki nilai sebesar 33.622,87, serta centroid data untuk kelompok dengan penyebaran yang lebih rendah, yang memiliki nilai 2.352,402.

Berdasarkan penelitian terdahulu yang telah dilakukan, hal ini dapat menambahkan pemahaman yang lebih mendalam tentang sifat pengelompokan distribusi listrik dan cara mempersingkat waktu komputerisasi menggunakan algoritma Mini Batch K-Means. Penelitian ini juga akan diolah berdasarkan wilayah geografis untuk mengidentifikasi perbedaan dibandingkan dengan penelitian sebelumnya [25]-[26]-[27].

2. METODOLOGI PENELITIAN

Metode penelitian mengacu pada prosedur dan teknik sistematis yang digunakan oleh peneliti untuk melakukan penelidikannya. Sasarannya adalah menata kemajuan penelitian secara metodis dan terorganisir, serta menjamin hasil yang dicapai selaras dengan tujuan yang diharapkan. Gambar 1 menggambarkan urutan proses yang harus dipatuhi ketika melakukan penelitian.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Tujuan tahap pengumpulan data adalah untuk memperoleh informasi atau mengumpulkan data yang relevan dengan penelitian. Data yang digunakan dalam penelitian ini merupakan data yang sudah ada sebelumnya yang bersumber dari Statistik PLN, khususnya mengenai sebaran listrik di setiap provinsi di Indonesia. Negara Indonesia terdiri dari 35 provinsi. Data yang dipakai meliputi rentang waktu dari tahun 2014 hingga 2022. Data distribusi listrik ini diukur menggunakan satuan yang sama, yaitu Giga Watt Hour (GWH) [7]. Sumber data berasal dari situs web <https://web.pln.co.id/>. Penelitian ini melakukan analisis statistik deskriptif terhadap data distribusi listrik di semua provinsi di Indonesia.

2.2. Pre-Processing

Tahap *pre-processing* adalah serangkaian aktivitas yang dilakukan sebelum memulai pengolahan data. Tahap ini menyusun langkah-langkah penting yang harus dilakukan dengan segera, bertujuan untuk menghemat waktu dan usaha. Tahapan *Knowledge Discovery* melibatkan serangkaian langkah yang harus dilakukan pada data agar dapat dimanfaatkan untuk proses *data mining*. *Preprocessing data mining* adalah proses mengolah data mentah menjadi data yang siap digunakan, dan salah satu tahapnya adalah *Data Transformation*.

2.3. Pengolahan Data

Tahapan ini dijalankan setelah proses preprocessing selesai, sebelum melanjutkan ke tahap *data mining*. pengelompokan data pendistribusian di seluruh provinsi indonesia menggunakan metode algoritma *Mini Batch K-Means*. Yang akan dijelaskan menggunakan pseudocode algorithm 1 sebagai berikut:

Algorithm 1 Mini Batch K-means

Input: X is the similarity matrix $\{x_1, x_2, \dots, x_n\}$, in which n is the number of input values, mini-batch size (b), iterations (t), cluster number (k), the total number of features (d),
 d_{nom} feature are nominal features, d_{ord} the feature is ordinal features. Output: Number of Clusters, $C = \{C_1, C_2, \dots, C_k\}$

1. Start
2. Establish the initialization of k cluster nodes, $\mu = \{\mu_{c_1}, \mu_{c_2}, \dots, \mu_{c_k}\}$
3. Create each cluster, $C_i = \theta (1 \leq i \leq k)$
4. Initialize the number of clusters with data., $N_c = (1 \leq i \leq k)$
5. for $p = 1$ to t do:
6. $M = \{x_m \mid 1 \leq m \leq b\}$ # M is the batch dataset, and x_m is a random sample from X
7. for $m = 1, 2, \dots, b$ do
8. $\mu_{c_1}(x_m) = \frac{1}{|C_1|} \sum_{x_m \in C_1} x_m (x_m \in M)$
9. end for
10. For $m = 1, 2, \dots, b$ do
11. $\mu_{c_1} = \mu_{c_1}(x_m)$
12. $N_{c_1} = N_{c_1} + 1$
13. $\rho = \frac{1}{N_{c_1}}$
14. $\mu_{c_1} = (1 - \rho) \mu_{c_1} + \rho x_m$
15. end for
- end for [28]

2.4. Pengujian

Pada saat ini, proses pengujian akan dilakukan terhadap hasil yang dicapai sebelumnya. proses data mining menggunakan *python*. Hal ini berguna untuk melihat apakah hasil yang didapatkan sesuai dan juga untuk mengetahui apakah masih ada kesalahan dalam proses data mining [13]. Pengujian ini akan dilakukan dengan menggunakan *Silhouette Coefficient*. Pengujian dilakukan untuk mengukur seberapa baik objek dalam suatu cluster berdekatan satu sama lain dan seberapa terpisah dari cluster lain.

Rentang nilai adalah dari -1 hingga 1. Ketika koefisien siluet mendekati 1, pengelompokan data dalam satu cluster semakin baik. Namun, jika mendekati -1, pengelompokan data dalam suatu cluster menjadi semakin buruk. Metode ini mengkuantifikasi tingkat kemiripan antara setiap titik dan cluster yang dimilikinya, relatif terhadap cluster lainnya. Tujuannya adalah untuk memberikan gambaran grafis tentang keakuratan klasifikasi setiap objek [25]. Untuk memberikan informasi tentang seberapa baik hasil pengelompokan pada proses pengelompokan, silhouette dapat dihitung untuk setiap cluster dan juga untuk keseluruhan cluster dari hasil kerja suatu algoritma pengelompokan. Adapun langkah langkah persamaan mencari *Silhouette Coefficient* adalah sebagai berikut:

1. Hitung nilai $a(i)$ dengan menghitung rata-rata jarak titik tersebut dengan semua titik lain dalam cluster yang sama.
2. Hitung nilai $b(i)$ dengan menghitung rata-rata jarak titik tersebut dengan semua titik dalam cluster terdekat dengan cluster tempatnya berada.
3. Gunakan rumus di bawah ini untuk menghitung koefisien siluet atau nilai $s(i)$ untuk titik ke- i .

$$s(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))} \quad (1)$$

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini menggunakan sumber sekunder, yaitu laporan yang dikeluarkan Statistik PLN. Laporan ini mencakup informasi tentang distribusi listrik di setiap provinsi di Indonesia. Data yang disertakan dalam analisis ini mencakup periode tahunan dari 2014 hingga 2022. Ada tujuh variabel yang digunakan untuk mengkategorikan data, yang mencakup satuan konsumsi listrik yang berbeda-beda, seperti unit PLN di setiap provinsi, rumah tangga, industri, dunia usaha, keperluan sosial, gedung perkantoran pemerintah, dan jalan raya umum. Algoritme akan menggunakan atribut ini selama prosedur pengelompokan menggunakan Mini Batch K-means. Data yang ditampilkan pada Tabel 1 hanya Salah satu data distribusi listrik.

Tabel 1. Data Distribusi 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
....
35	Dist. Jakarta Raya	14,824.00	4,140.34	12,539.23	1,482.92	1,396.61	195.20

3.2. Pre-Processing Data

Tahap Pre-Processing dalam analisis data melibatkan serangkaian langkah untuk mempersiapkan data mentah sebelum masuk ke dalam proses analisis. Salah satu proses krusial dalam tahap ini adalah tahap transformasi, di mana data diubah menjadi format yang lebih sesuai untuk analisis lanjutan. Tahap transformasi data dilakukan dengan menjumlahkan semua konsumsi listrik dari tahun 2014 sampai tahun 2022. Kemudian hasil akumulasi data untuk 35 propinsi tersebut (diberikan label untuk singkatan nama propinsi atau satuan PLN), dikelompokkan berdasarkan letak geografis dan kedekatan wilayahnya, menjadi 4 kelompok, yaitu:

1. Sumatera, meliputi: Aceh, Sumut, Sumbar, Riau, Kepri, Batam, Jambi, Bengkulu, Sumsel, Babel, dan Lampung (11 wilayah).
2. Jawa-Bali, meliputi: Banten, DKI/Tangerang, Jabar, Jateng, Jatim, DIY, Bali, NTB dan NTT (9 wilayah)
3. Kalimantan-Sulawesi, meliputi: Kalbar, Kalteng, Kaltim/Kaltara, Kalsel, Sulut, Gorontalo, Tarakan, Sulbar, Sulsel, Sulteng, Sultra, Maluku, dan Maluku (13 wilayah)
4. Papua.

Hasil proses transformasi data disajikan dalam bentuk Tabel 2-5.

Tabel 2. Data Akumulasi Distribusi Listrik Tahun 2014 – 2022 Regional Sumatra

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

No	Satuan PLN/ Provinsi	Rumah Tangga	Industri	Bisnis	Sosial	Gedung Kantor Pemerintah	Jalan Umum
3	Wilayah Sumatera Barat	14,938.30	8056.82	4514.05	1311.53	621.73	839.32
....
11	Distribusi Lampung	23,329.31	8,254.99	4,789.97	1,380.17	557.4	931.87

Tabel 3. Data Akumulasi Distribusi Listrik Tahun 2014 – 2022 Regional Jawa-Bali

No	Satuan PLN/ Provinsi	Rumah Tangga	Industri	Bisnis	Sosial	Gedung Kantor Pemerintah	Jalan Umum
1	Wilayah Banten	40,207.45	111,446.98	21,891.22	2,311.47	1,161.61	673.2
2	Wilayah Jawa Barat	166,559.08	20,5651.98	50,909.48	9,347.15	3722	3,105.9
3	Distribusi Jakarta Raya	125,995.80	5,1211.27	108,466.03	11,719.59	11,799.27	2,684.04
....
9	Wilayah NTT	5,258.83	350	1,977.35	469.83	368.38	139.91

Tabel 4. Data Akumulasi Distribusi Listrik Tahun 2014 – 2022 Regional Kalimantan-Sulawesi

No	Satuan PLN/ Provinsi	Rumah Tangga	Industri	Bisnis	Sosial	Gedung Kantor Pemerintah	Jalan Umum
1	Wilayah Kal. Barat	13,888.16	1,438.1	4,653.8	867.49	659.33	366.74
2	Wilayah Kal. Tengah	7,398.78	751.14	2,361.72	439.21	484.59	179.6
3	Wilayah Kal. Selatan	14,416.14	2,905.6	4,284.3	923.29	609.54	559.59
....
13	Wilayah Maluku Utara	2,746.39	41.18	745.18	174.61	271.85	79

Tabel 5. Data Akumulasi Distribusi Listrik Tahun 2014 – 2022 Regional Papua

No	Satuan PLN/ Provinsi	Rumah Tangga	Industri	Bisnis	Sosial	Gedung Kantor Pemerintah	Jalan Umum
1	Wilayah Papua Barat	2,828.96	76.33	1,228.28	234.52	322.77	60.68
2	Wilayah Papua	5,098.88	68.38	2,461.48	442.11	634.99	97.32

Berdasarkan data pada Tabel 5, terdapat hanya dua wilayah distribusi listrik di wilayah Papua yang tidak perlu dimasukkan dalam proses clustering karena jumlah wilayah yang terbatas tidak memungkinkan analisis yang bermakna.

3.3. Penerapan algoritma Mini Batch K-means

Penerapan algoritma Mini Batch K-means Clustering adalah langkah penting dalam analisis data yang berpusat pada pengelompokan data yang terkait. Ini merupakan tahap krusial dalam proses analisis, di mana fokus utamanya adalah pada pengelompokan data terkait distribusi listrik. Dalam penerapan clustering algoritma Mini Batch K-means ada beberapa tahapan yang harus dijalankan. Data yang digunakan merupakan data hasil inialisasi menggunakan Mini Batch K-means.

Bagian berikut ini adalah simulasi perhitungan pembentukan 2 cluster dari data yang tersedia, sesuai dengan algoritma mini batch K-means seperti dalam pseudocode Algorithm 1. Untuk pembentukan 3 cluster menyesuaikan. Input data merupakan proses memasukkan data distribusi listrik yang akan di cluster. Data distribusi listrik yang akan di cluster dapat ditemukan pada bagian Tabel 2-4 yang sudah diakumulasikan. Centroid Data Menentukan berapa banyak kelompok yang akan dibentuk untuk setiap regional. Dalam algoritma ini, akan dibentuk dua kelompok (cluster) untuk mendapatkan nilai centroid data. Berdasarkan Tabel 6-7 merupakan nilai data centroid awal untuk setiap regional.

Tabel 6. Data Centroid awal Cluster 1

No	Regional	Rumah Tangga	Industri	Bisnis	Sosial	Gedung Kantor Pemerintah	Jalan Umum
1	Jawa-Bali	0.847	0.577	0.657	0.862	0.602	0.822
2	Kalimantan-Sulawesi	0.368	0.345	0.387	0.379	0.390	0.291
3	Sumatera	0.643	0.503	0.471	0.647	0.700	0.509

Tabel 7. Data Centroid awal Cluster 2

No	Regional	Rumah Tangga	Industri	Bisnis	Sosial	Gedung Kantor Pemerintah	Jalan Umum
1	Jawa-Bali	1.127	0.858	0.894	1.139	0.815	1.044
2	Kalimantan-Sulawesi	1.729	1.337	1.891	1.781	2.035	1.313

No	Regional	Rumah Tangga	Industri	Bisnis	Sosial	Gedung Kantor Pemerintah	Jalan Umum
3	Sumatera	0.842	0.651	0.661	0.888	0.954	0.731

Clustering Data adalah dengan memilih sebagian kecil data (*Mini-Batch*) secara acak dari dataset. Ukuran *Mini Batch* ini biasanya lebih kecil dari ukuran dataset keseluruhan. Hasil dari nilai centroid setiap variabel setelah proses clustering menggunakan metode *Mini Batch* akan disajikan dalam Tabel 8-9.

Tabel 8. Pusat Centroid Cluster 1

No	Regional	Rumah Tangga	Industri	Bisnis	Sosial	Gedung Kantor Pemerintah	Jalan Umum
1	Jawa-Bali	17962.996	23348.446	10854.044	1364.722	694.792	493.914
2	Kalimantan-Sulawesi	6063.939	808.649	1924.185	402.01	365.53	229.589
3	Sumatera	24919.491	9373.798	7344.551	1823.348	830.436	1487.768

Tabel 9. Pusat Centroid Cluster 2

No	Regional	Rumah Tangga	Industri	Bisnis	Sosial	Gedung Kantor Pemerintah	Jalan Umum
1	Jawa-Bali	128951.412	115210.93	56847.1225	9981.53	5315.34	3849.2975
2	Kalimantan-Sulawesi	22095.955	6943.355	8922.325	1777.49	1309.61	811.13
3	Sumatera	6242.628	1914.998	3046.09	391.478	299.73	153.13

Untuk setiap batch data yang diproses, evaluasi jarak antara setiap titik data dalam batch tersebut dengan pusat centroid yang ada. Kemudian, alokasikan masing-masing titik setiap variabel data ke dalam cluster yang memiliki centroid terdekat berdasarkan aturan rumus jarak *Euclidean*. Proses algoritma Mini Batch K-means untuk pencarian hasil perhitungan jarak setiap regional yang dapat digambarkan pada Tabel 10-12.

Tabel 10. Hasil Perhitungan Jarak Euclidean Regional Jawa-Bali

No	Provinsi	C1	C2
1	Wilayah Banten	0.576	4.769
2	Wilayah Jawa Barat	0.467	4.602
3	Distribusi Jakarta Raya dan Tangerang	0.389	4.366
4	Wilayah Jawa Tengah	0.479	4.177
...
9	Wilayah NTT	4.599	1.520

Tabel 11. Hasil Perhitungan Jarak Euclidean Regional Kalimantan-Sulawesi

No	Provinsi	C1	C2
1	Wilayah Kal. Barat	1.639	6.654
2	Wilayah Kal. Tengah	1.145	6.176
3	Wilayah Kal. Selatan	0.926	5.940
4	Wilayah Kal. Timur dan Utara	0.806	5.721
...
13	Wilayah Maluku Utara	7.195	2.633

Tabel 12. Hasil Perhitungan Jarak Euclidean Regional Sumatera

No	Provinsi	C1	C2
1	Wilayah Aceh	3.663	0.388
2	Wilayah Sumatera Utara	3.811	0.580
3	Wilayah Sumatera Barat	3.621	0.377
...
10	Wilayah Riau	2.853	1.391

Proses hasil Euclidean Mini Batch Kmeans akan ditentukan data yang masuk kedalam suatu cluster. Data yang akan menjadi bagian dari sebuah kelompok adalah data yang paling dekat dengan titik pusat Clusternya, seperti yang ditunjukkan dalam Tabel 13-15.

Tabel 13. Hasil Cluster Regional Jawa-Bali

No	Provinsi	C1	C2
1	Wilayah NTT	1	
2	Wilayah NTB	1	
3	Distribusi D.I Yogyakarta	1	
4	Distribusi Bali	1	
5	Wilayah Banten	1	
6	Wilayah Jawa Tengah		1
7	Distribusi Jawa Timur		1
8	Dist. Jarkarta Raya dan Tangerang		1
9	Wilayah Jawa barat		1

Tabel 14. Hasil Cluster Regional Kalimantan-Sulawesi

No	Provinsi	C1	C2
1	PT PLN Tarakan	1	
2	Wilayah Sulawesi Barat	1	
3	Wilayah Maluku Utara	1	
4	Wilayah Maluku	1	
5	Wilayah Gorontalo	1	
6	Wilayah Sulawesi Tenggara	1	
7	Wilayah Sulawesi Tengah	1	
8	Wilayah Kalimantan Tengah	1	
9	Wilayah Sulawesi Utara	1	
10	Wilayah Kalimantan Barat	1	
11	Wilayah Kalimantan Selatan		
12	Wilayah Kal. Timur dan Utara		1
13	Wilayah Sulawesi Selatan		1

Tabel 15. Hasil Cluster Regional Kalimantan-Sulawesi

No	Provinsi	C1	C2
1	Wilayah Kepulauan Riau		1
2	Wilayah Bengkulu		1
3	Wilayah Bangka Belitung		1
4	PT PLN Batam		1
5	Wilayah Jambi		1
6	Wilayah Aceh	1	
7	Wilayah Sumatera Barat	1	
8	Wilayah Riau	1	
9	Distribusi Lampung	1	
10	Wilayah Sumatera Selatan	1	
11	Wilayah Sumatera Utara	1	

Berdasarkan hasil perhitungan pada Tabel 13-15 untuk wilayah Regional Jawa-Bali, Kalimantan-Sulawesi, dan Sumatera, dihasilkan cluster distribusi energi dibagi menjadi dua kategori: Cluster 1 yang mencakup wilayah dengan distribusi listrik tingkat tinggi dan Cluster 2 yang mencakup wilayah dengan distribusi listrik tingkat rendah. Di Regional Jawa-Bali, wilayah dengan distribusi listrik tingkat tinggi (Cluster 1) meliputi NTT, NTB, D.I Yogyakarta, Bali, dan Banten, sementara wilayah dengan distribusi listrik tingkat rendah (Cluster 2) meliputi Jawa Tengah, Jawa Timur, Jakarta Raya dan Tangerang, serta Jawa Barat.

Di Regional Kalimantan-Sulawesi, wilayah dengan distribusi listrik tingkat tinggi (Cluster 1) meliputi PT PLN Tarakan, Sulawesi Barat, Maluku Utara, Maluku, Gorontalo, Sulawesi Tenggara, Sulawesi Tengah, Kalimantan Tengah, Sulawesi Utara, dan Kalimantan Barat, sedangkan wilayah dengan distribusi listrik tingkat rendah (Cluster 2) meliputi Kalimantan Selatan, Kalimantan Timur dan Utara, serta Sulawesi Selatan.

Di Regional Sumatera, wilayah dengan distribusi listrik tingkat tinggi (Cluster 1) meliputi Aceh, Sumatera Barat, Riau, Lampung, Sumatera Selatan, dan Sumatera Utara, sementara wilayah dengan distribusi listrik tingkat rendah (Cluster 2) meliputi Kepulauan Riau, Bengkulu, Bangka Belitung, Batam, dan Jambi. Pengelompokan ini menunjukkan variasi dalam kebutuhan dan infrastruktur listrik di masing-masing wilayah, yang dapat dijadikan dasar untuk menetapkan prioritas dalam pengembangan infrastruktur listrik yang lebih efisien dan merata sesuai dengan kebutuhan regional.

Oleh karena itu, studi ini menawarkan wawasan penting bagi individu yang bertugas dalam perencanaan strategis dan implementasi distribusi Listrik di Indonesia. Hal ini memperhitungkan variabel-variabel sekitar yang mempengaruhi alokasi listrik yang efektif dan adil.

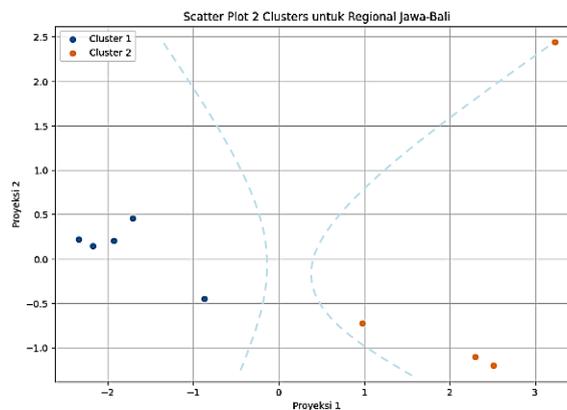
3.4. Pengujian Silhouette Score

Pada penelitian ini akan dilakukan pengujian 2 cluster, 3 cluster, dan 4 cluster. Hasil pengujian *Silhouette Score* ini menggunakan persamaan (1) untuk setiap wilayah dalam masing-masing regional. Dalam pembahasan ini akan ditunjukkan hasil evaluasi dari *Silhouette Score* untuk proses pengelompokan data di setiap wilayah regional. Hasil dari pengujian ini ditunjukkan pada Tabel 16.

Tabel 16. Hasil Pengujian *Silhouette Score*

No	Regional	Nilai Silhouette Score untuk jumlah cluster		
		N=2	N=3	N=4
1	Jawa-Bali	0.573	0.411	0.451
2	Kalimantan-Sulawesi	0.625	0.484	0.419
3	Sumatera	0.443	0.512	0.471

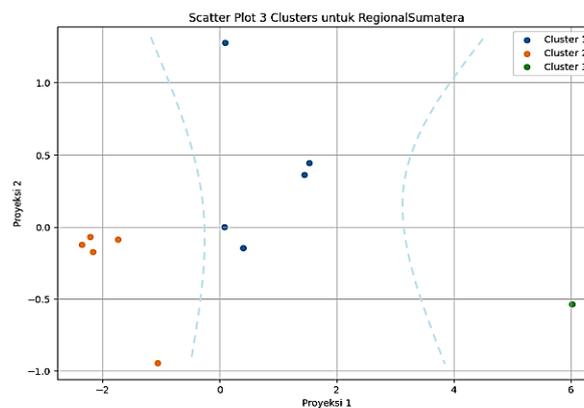
Hasil cluster yang terbentuk disetiap wilayah regional memiliki nilai yang berbeda-beda. Dengan perbedaan hasil penelitian ini akan mengambil rekomendasi jumlah cluster nilai *Silhouette Score* terbaik. Pada Tabel 16 Regional Jawa-Bali 2 Cluter memiliki nilai 0,573 , Regional Kalimantan-Sulawesi di 2 Cluster memiliki nilai 0,625 dan Regional Sumatera 3 Cluster memiliki nilai 0,512. Gambar 2-4 berikut merupakan hasil scatter plot dari cluster yang memiliki nilai *Silhouette Score* Tertinggi.



Gambar 2. Scatter Plot Cluster Regional Jawa-Bali.

Tabel 17. Hasil 2 Cluster Regional Jawa-Bali.

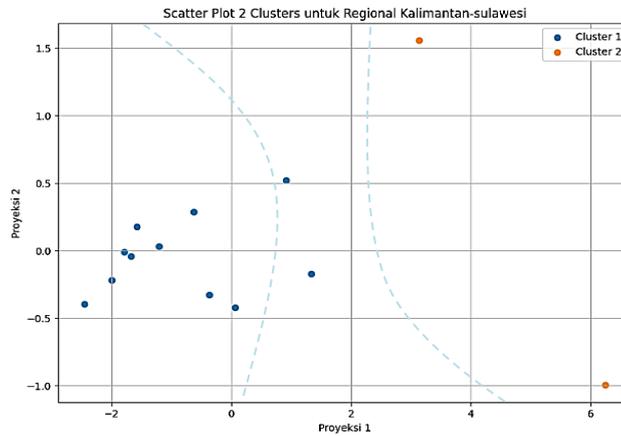
C1	C2
NTT	Jateng
NTB	Jatim
D.I.Y	DKI dan TNG
Bali	Jabar
Banten	



Gambar 3. Scatter Plot 3 Cluster Regional Sumatera

Tabel 18. Hasil 3 Cluster Regional Sumatera

C1	C2	C3
Aceh	Kepri	Sumut
Sumbar	Bengkulu	
Riau	Babel	
Lampung	PLN Batam	
SumSel	Jambi	



Gambar 4. Scatter Plot 2 Cluster Regional Kal-Sul.

Tabel 19. Hasil 2 Cluster Regional Kal-Sul

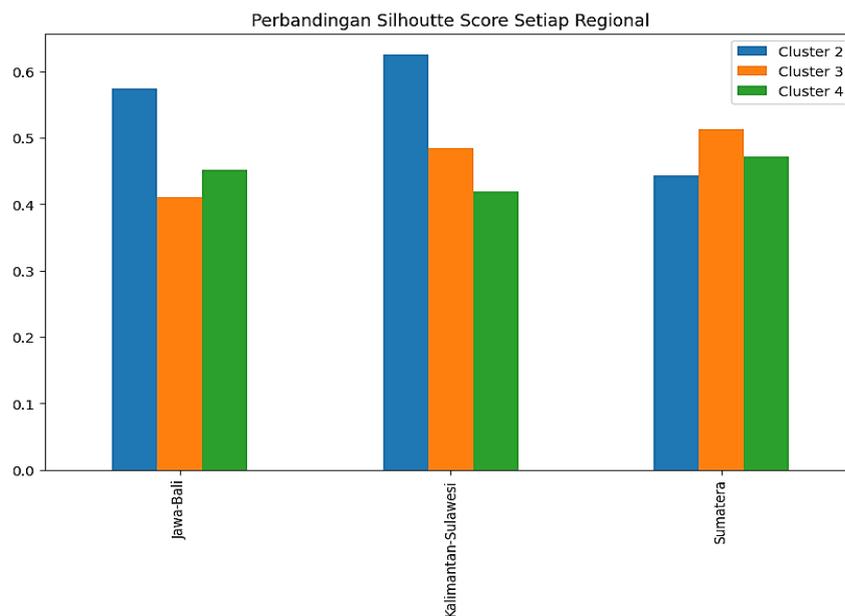
C1	C2
Tarakan	Kaltim dan Kaltara
Sulbar	Sulsel
Malut	
Maluku	
Gorontalo	
Sultang	
Sulteng	
Kalteng	
Sulut	
Kalbar	
Kalsel	

Tabel 17-19 memperlihatkan anggota cluster yang terbentuk dari hasil proses clustering dengan metode ini, untuk ketiga wilayah regional berdasarkan nilai Silhouette tertinggi. Untuk wilayah regional Sumatera terbentuk 3 cluster, sedangkan wilayah Jawa-Bali dan Kalimantan-Sulawesi terbentuk 2 cluster. Hasil ini menunjukkan pula profil dan karakteristik dari anggota-anggota clusternya, misalnya pada wilayah Sumatera, Sumatera utara berada dalam cluster tersendiri, karena dari segi pertumbuhan ekonomi dan kepadatan penduduk, lebih tinggi dari propinsi lainnya, sehingga konsumsi listrik untuk berbagai bidang pun berbeda. Sedangkan untuk wilayah Jawa-Bali, propinsi dengan jumlah penduduk terpadat, area terluas dan industri yang banyak berkumpul dalam satu cluster, yaitu DKI/ TNG, Jabar, Jateng, dan Jatim. Di wilayah Kalimantan-Sulawesi juga sama, Kaltim/Kaltara dan Sulsel menempati cluster yang sama, dengan profil yang hampir serupa dengan cluster "padat" di wilayah regional lain.

Berdasarkan hasil pengujian Silhouette Score untuk pengelompokan tiga regional (Jawa-Bali, Kalimantan-Sulawesi, dan Sumatera), dapat disimpulkan bahwa pengelompokan menggunakan dua cluster (Cluster 2) menunjukkan hasil yang paling efektif. Silhouette Score, yang mengukur seberapa mirip suatu objek dengan klusternya sendiri dibandingkan dengan kluster lain, menunjukkan nilai tertinggi untuk Cluster 2 di ketiga regional tersebut: Jawa-Bali dengan skor 0.573, Kalimantan-Sulawesi dengan skor 0.625, dan Sumatera dengan skor 0.443. Sebaliknya, Cluster 3 dan Cluster 4 menunjukkan nilai Silhouette Score yang lebih rendah di semua regional, dengan Cluster 3 memiliki nilai tertinggi 0.512 di Sumatera dan Cluster 4 dengan nilai tertinggi 0.471 di Sumatera. Oleh karena itu, pengelompokan menggunakan dua cluster lebih efektif dan memberikan hasil yang lebih baik dalam hal kedekatan internal dan keterpisahan antar cluster. Pengelompokan ini dapat digunakan sebagai dasar untuk merancang strategi distribusi listrik yang lebih efisien dan tepat sasaran di setiap regional, memastikan pengelompokan yang lebih jelas dan terdefinisi dengan baik. Dapat dilihat Gambar 5 Hasil dari perbandingan *Silhouette Score* setiap Cluster.

Dari hasil cluster menggunakan metode Mini Batch Kmeans terlihat bahwa untuk 2 cluster regional jawa-bali, Provinsinya memiliki cluster yang sama dengan Hasil [29] yang menggunakan metode DBSCAN, di mana untuk jawa-bali juga terbentuk 2 cluster dengan anggota yang sama. bila dibentuk 3 cluster regional sumatra (Walaupun bukan silhouette score terbaik) hasil cluster lebih relevan menggambarkan kondisi real, yang sesuai dengan keadaan demografi, profil industri, dan rumah tangga di area tersebut. Pada cluster 3 juga memiliki hasil yang sama pada wilayah provinsinya [30] yang menggunakan algoritma MeanShift. Hal ini dapat dipelajari lebih lanjut dari data statistik pada situs resmi BPS¹.

Hasil uji Silhouette Score untuk setiap cluster menawarkan penilaian komprehensif terhadap kualitas pengelompokan dalam konteks distribusi listrik di berbagai regional. Hasil-hasil ini dapat menjadi landasan untuk mengambil keputusan yang tepat dalam perancangan dan pembangunan infrastruktur Listrik masa depan.



Gambar 3. Perbandingan Nilai *Silhouette Score*

4. KESIMPULAN

Ia berhasil menerapkan metode Mini Batch K-Means dalam penelitiannya untuk mengkategorikan data distribusi energi ke dalam variabel tertentu seperti perumahan, industri, bisnis, sosial, gedung pemerintah, dan jalan raya umum. Hasil pengujian menunjukkan bahwa terdapat 3 cluster yang signifikan dalam data peneliti, masing-masing ditandai dengan Silhouette Score yang berbeda. Cluster dengan Silhouette Score tertinggi secara konsisten adalah Cluster 2, menunjukkan kohesi yang tinggi dengan nilai 0.625 di regional Kalimantan-Sulawesi yang diamati dalam cluster ini. Sementara itu, Cluster 3, dan 4 memiliki Silhouette Score yang lebih rendah, menunjukkan tingkat kohesi yang lebih rendah atau kekhasan yang lebih rendah dengan nilai 0,421 pada regional Jawa-Bali dan 0,451 di regional Kalimantan-Sulawesi dalam pola distribusi listrik mereka. Dari temuan penelitian, disarankan untuk menganalisis lebih lanjut Cluster 2 yang menunjukkan kohesi tinggi dalam distribusi listrik. Cluster 3, dan 4 meskipun memiliki nilai Silhouette Score yang lebih rendah, masih dapat dimanfaatkan untuk pengoptimalan infrastruktur dan layanan terkait listrik. Batasan yang dimiliki hanya bisa dilakukan pengelompokan sampai 4 cluster. Secara keseluruhan, penelitian ini menegaskan bahwa algoritma Mini Batch K-means dapat digunakan secara efektif untuk pengelompokan data distribusi listrik dengan hasil yang memuaskan dalam pemisahan klaster yang berbeda-beda.

REFERENSI

- [1] E. Manik, "Data Mining Pengelompokan Pemakaian Listrik Terboros Menggunakan Metode Clustering (Studi Kasus PLN Area Binjai)," *Inf. Syst. Dev.*, vol. 3, no. 2, pp. 116–119, 2018, [Online]. Available: <https://ejournal.medan.uph.edu/index.php/isd/article/view/211> (10 Juni 2020)
- [2] T. T. Gultom, "Pemenuhan sumber tenaga listrik di Indonesia," *J. Ilm. Res. Sains*, vol. 3, no. 1, pp. 130–139, 2017.
- [3] R. Aryani, M. Nasrun, C. Setianingsih, and M. A. Murti, "Clustering Data in Power Management System Using k-Means Clustering Algorithm," in *2019 IEEE Asia Pacific Conference on Wireless and*

¹ <https://www.bps.go.id/id>

- Mobile (APWiMob)*, 2019, pp. 164–170. doi: 10.1109/APWiMob48441.2019.8964143.
- [4] J. Mangundap, S. Silimang, and Tumaliang Hans, “Analisa Rugi-Rugi Daya Jaringan Distribusi Di PT. PLN (Persero) Area Manado 2017,” *J. Tek. Elektro dan Komput.*, vol. Vol. 7, no. NO. 3, pp. 1. – 5, 2018.
 - [5] H. Hendrocahyo and L. Kurniawati, “Understanding the Financial Performance of PT PLN (Persero): A Narrative on State-Owned Enterprise (SOE) with a Mandate of Electricity in Indonesia,” *Binus Bus. Rev.*, vol. 13, no. 3, pp. 241–258, 2022, doi: 10.21512/bbr.v13i3.7883.
 - [6] K. Wang, R. Yang, C. Liu, T. Samarasinghalage, and Y. Zang, “Extracting Electricity Patterns from High-dimensional Data: A comparison of K-Means and DBSCAN algorithms,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1101, no. 2, 2022, doi: 10.1088/1755-1315/1101/2/022007.
 - [7] D. M. P. P. L. T. M. M. C. (Studi K. P. A. Binjai), “Pengelompokan Data Time Series Pada Distribusi Listrik,” vol. 11, no. 5, pp. 735–742, 2022.
 - [8] T. Cerquitelli *et al.*, “Discovering electricity consumption over time for residential consumers through cluster analysis,” *2018 14th Int. Conf. Dev. Appl. Syst. DAS 2018 - Proc.*, pp. 164–169, 2018, doi: 10.1109/DAAS.2018.8396090.
 - [9] “Geofluids - 2018 - Su - Sedimentary Environment Analysis by Grain-Size Data Based on Mini Batch K-Means Algorithm.pdf.”
 - [10] C. Astria, A. P. Windarto, A. Wanto, and ..., “Metode K-Means Pada Pengelompokan Wilayah Pendistribusian Listrik,” ... *Nas. Sains dan ...*, 2019, [Online]. Available: <http://prosiding.seminar-id.com/index.php/sensasi/article/view/317>
 - [11] “adoc.pub transmisi-dan-distribusi-tenaga-listrik”.
 - [12] T. Wang, C. Ren, Y. Luo, and J. Tian, “NS-DBSCAN: A density-based clustering algorithm in network space,” *ISPRS Int. J. Geo-Information*, vol. 8, no. 5, May 2019, doi: 10.3390/ijgi8050218.
 - [13] R. D. Herdiansyah, F. Sains, D. A. N. Teknologi, U. Islam, N. Sultan, and S. Kasim, “Penerapan Algoritma K-Means Clustering Pada Penentuan Tingkat Pemakaian Listrik (Studi Kasus : PLTBg Rantau Sakti),” 2023.
 - [14] S. Saputra, *Analisis Kelayakan Penerima Bantuan Covid-19 Menggunakan Metode K-Means pada Kecamatan Sagulung Kota Batam*. repository.upbatam.ac.id, 2021. [Online]. Available: <http://repository.upbatam.ac.id/1052/>
 - [15] K. Peng, V. C. M. Leung, and Q. Huang, “Clustering Approach Based on Mini Batch Kmeans for Intrusion Detection System over Big Data,” *IEEE Access*, vol. 6, pp. 11897–11906, 2018, doi: 10.1109/ACCESS.2018.2810267.
 - [16] M. A. Ahmed, H. Baharin, and P. N. E. Nohuddin, “Mini-Batch k-Means versus k-Means to Cluster English Tafseer Text: View of Al-Baqarah Chapter,” *J. Quranic Sci. Res.*, vol. 02, no. 02, pp. 48–53, 2021, doi: 10.30880/jqsr.2021.02.02.006.
 - [17] C. Astria, D. Hartama, A. P. Windarto, and ..., “Pengembangan Metode Datamining K-Medoid Pada Kasus Distribusi Listrik di Indonesia,” ... *Komput. & Sains ...*, 2020, [Online]. Available: <http://prosiding.seminar-id.com/index.php/sainteks/article/view/447>
 - [18] F. P. Rachman, H. Santoso, and A. Djajadi, “Machine Learning Mini Batch K-means and Business Intelligence Utilization for Credit Card Customer Segmentation,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 10, pp. 218–227, 2021, doi: 10.14569/IJACSA.2021.0121024.
 - [19] N. M. Mahfuz, M. Yusoff, M. S. Nordin, and Z. Ahmad, “Firefly Algorithm with Mini Batch K-Means Entropy Measure for Clustering Heterogeneous Categorical Timber Data,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 7, pp. 461–468, 2022, doi: 10.14569/IJACSA.2022.0130756.
 - [20] M. M. Chavan, A. Patil, L. Dalvi, and A. Patil, “Mini batch k-means clustering on large dataset,” *Int. J. Sci. Eng. Technol. Res.*, vol. 04, no. 07, pp. 1356–1358, 2015, [Online]. Available: www.ijsetr.com
 - [21] C. Science and C. Science, “Mini-Batch K-Means Clustering Using Map-Reduce in Hadoop,” vol. 2, no. 2, pp. 336–342, 2014.
 - [22] S. R. Fitriyani and H. Murfi, “The K-means with mini batch algorithm for topics detection on online news,” *2016 4th Int. Conf. Inf. Commun. Technol. ICoICT 2016*, vol. 4, no. c, 2016, doi: 10.1109/ICoICT.2016.7571914.
 - [23] A. Susanto and R. Roestam, “Segementasi Gambar pada Dataset MNIST dengan Optimasi Mini Batch dan K-means++ pada Algoritma K-means,” *J. Sist. dan Teknol. Inf.*, vol. 11, no. 2, p. 309, Jul. 2023, doi: 10.26418/justin.v11i2.55503.
 - [24] C. Astria, A. P. Windarto, A. Wanto, and ..., “Metode K-Means Pada Pengelompokan Wilayah Pendistribusian Listrik,” ... *Nas. Sains dan ...*, 2019, [Online]. Available: <http://prosiding.seminar-id.com/index.php/sensasi/article/view/317>
 - [25] F. T. Elektro, M. Ary, M. Fakultas, T. Elektro, and C. Setianingsih, “Pengelompokan Data Penggunaan Energi Listrik menggunakan Algoritma Mini Batch K-Means clustering.”
 - [26] T. Wahyuningrum, S. Khomsah, S. Suyanto, S. Meliana, P. E. Yunanto, and W. F. Al Maki, “Improving Clustering Method Performance Using K-Means, Mini Batch K-Means, BIRCH and Spectral,” *2021*

- 4th Int. Semin. Res. Inf. Technol. Intell. Syst. ISRITI 2021*, pp. 206–210, 2021, doi: 10.1109/ISRITI54043.2021.9702823.
- [27] S. C. Hicks, R. Liu, Y. Ni, E. Purdom, and D. Risso, “Mbkmeans: Fast clustering for single cell data using mini-batch k-means,” *PLoS Comput. Biol.*, vol. 17, no. 1, pp. 1–18, 2021, doi: 10.1371/JOURNAL.PCBI.1008625.
- [28] N. M. Mahfuz, M. Yusoff, and Z. Idrus, “Clustering heterogeneous categorical data using enhanced mini batch K-means with entropy distance measure,” *Int. J. Electr. Comput. Eng.*, vol. 13, no. 1, pp. 1048–1059, 2023, doi: 10.11591/ijece.v13i1.pp1048-1059.
- [29] M. Farid, F. Insani, S. Agustian, and L. Afriyanti, “Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Density Based Spatial Clustering Of Application With Noise (DBSCAN),” 2024.
- [30] R. Utari, F. Insani, S. Agustian, and L. Afriyanti, “Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Mean shift,” 2024.