

Pengelompokkan Data Penggunaan Energi Listrik menggunakan Algoritma Mini Batch K-Means clustering

1st Amanda Austin Herlambang

Fakultas Teknik Elektro
Universitas Telkom

Bandung, Indonesia

amandaastinh@student.telkomuni-
versity.ac.id

2nd Muhammad Ary Murti

Fakultas Teknik Elektro
Universitas Telkom

Bandung, Indonesia

arymurti@telkomuniversity.ac.id

3rd Casi Setianingsih

Fakultas Teknik Elektro
Universitas Telkom

Bandung, Indonesia

setiacasie@telkomuniversity.ac.id

Abstrak—Penggunaan energi listrik sudah menjadi kebutuhan yang pokok, sebagian besar pengguna menggunakan listrik tanpa menyadari besarnya listrik yang digunakan pada periode itu dapat membuat penggunaan listrik melonjak karena tidak ada kontrol penggunaan listrik. Clustering atau pengelompokkan data ini dibutuhkan untuk dapat mengetahui penggunaan energi listrik berlebih disuatu gedung. Pada penelitian ini dirancang sebuah sistem yang dapat memberikan informasi mengenai penggunaan listrik suatu gedung dengan menggunakan pengelompokkan data berbasis website. Pengelompokkan data ini menggunakan pembelajaran mesin unsupervised learning dengan algoritma Mini Batch K-Means dan terbagi menjadi tiga bagian pengelompokkan yaitu penggunaan energi listrik tinggi, normal dan rendah. Pengelompokkan data akan dilakukan untuk memonitoring penggunaan energi listrik perbulan, perhari dan pergedung. Dari hasil pengujian yang dilakukan menunjukkan bahwa nilai silhouette score clustering perhari menggunakan data Gedung N bernilai 0,62, perbulan data Gedung N sebesar 0,57, seluruh hasil tersebut termasuk ke dalam struktur baik, pertahun data Gedung N sebesar 0,73 termasuk ke dalam struktur kuat. Clustering menggunakan data dummy Gedung P dan Gedung O adalah sebesar 0,55 untuk perhari yang termasuk ke dalam struktur baik, perbulan sebesar 0,50 termasuk ke dalam struktur lemah dan pertahun sebesar 0,72 termasuk ke dalam struktur kuat.

Kata kunci— Energi listrik, clustering, mini batch k-means clustering, pengelompokkan data.

I. PENDAHULUAN

Penggunaan energi listrik di era industry 4.0 sudah menjadi kebutuhan yang pokok, energi listrik merupakan energi yang dibutuhkan untuk menunjang segala aktivitas manusia contohnya adalah pada suatu gedung. Seiring dengan meningkatnya penggunaan listrik tersebut, maka dibutuhkan pengelompokkan data penggunaan energi listrik untuk membantu pengguna

memonitoring penggunaan energi listrik secara nyata berdasarkan waktu agar dapat mengoptimalkan penggunaannya. Sebagian besar pengguna, menggunakan listrik tanpa menyadari besarnya listrik yang digunakan pada periode itu dapat membuat penggunaan listrik melonjak karena tidak ada kontrol penggunaan listrik [1]. Dengan menggunakan informasi dari pengelompokkan data, karyawan logistik dapat membuat kebijakan untuk melakukan penghematan listrik [1]. Adapun beberapa algoritma umum yang sering digunakan untuk pengelompokkan data atau *clustering* adalah DBSCAN [2], Fuzzy C-Means (FCM) [3], dan yang paling populer adalah K-Means [4] [5] yang dapat membuat suatu cluster yang dapat mengelompokkan data menjadi beberapa cluster berdasarkan kesamaan data. Mekanisme ini memungkinkan data yang memiliki kesamaan karakteristik dikelompokkan menjadi satu cluster dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam cluster lainnya. Proses pengelompokkan data pada algoritma K-Means adalah berdasarkan data numerik yang menjadi sumber datanya [4] Pada penelitian sebelumnya telah dilakukan clustering menggunakan algoritma K-Means dengan tingkat akurasi sebesar 83,3% [1]. Implementasi baru *clustering* penggunaan energi listrik rumah tangga menggunakan metode *clustering* Fuzzy Subtractive oleh Nurul Ramadhani Hikmiah, Riki Ruli A. Siregar, Budi Prayitno, Dine Tiara Kusuma, Novi Gusti Pahiyanti (2021) menghasilkan keluaran tingkatan pemakaian listrik pengguna tinggi, sedang dan rendah. Berdasarkan hasil pengujian menunjukkan nilai *Silhouette Coefficient* terbaik sebesar 0.8322535 [6]. Pada penelitian ini tidak hanya melakukan *clustering* penggunaan energi listrik menggunakan metode Mini Batch K-Means saja, tetapi juga mengelompokkan penggunaan energi listrik perhari, perbulan dan pertahun dan terbagi ke dalam kategori penggunaan energi rendah,

normal dan tinggi pada Gedung N, Gedung P dan Gedung O menggunakan metode Mini Batch K-Means clustering.

II. KAJIAN TEORI

A. Time Series Data

Time series merupakan deretan atau urutan waktu yang terjadi secara sekuensial dengan tujuan untuk mendapatkan kesimpulan data yang sudah dilakukan sebelumnya. Dimensi waktu dalam time series ini dapat berfungsi sebagai *feature* untuk memberikan banyak kegunaan dalam pemrosesan dan analisis data, kemudian dapat memperoleh *insight*, atau kesimpulan terhadap suatu rentetan data yang telah di amati sebelumnya [7]. Analisis pola waktu yang terjadi sangat penting dilakukan untuk prediksi menggunakan time series.

Time series dibagi menjadi dua yaitu Time Series Analysis dan Time Series Forecasting. Time series analysis berfokus pada pemahaman mengenai dataset. Time series forecasting memiliki tujuan untuk memprediksikan kejadian yang akan terjadi dimasa depan yang didapatkan melalui proses analisis dataset sebelumnya. Pada penelitian ini time series data yang digunakan adalah dalam bentuk *stationary*, data time series akan diubah ke suatu bentuk pola yang dikatakan

konstan selama deret waktu tertentu. Pencarian nilai *stationary* dapat dilakukan menggunakan rumus

differencing[8] yang dapat ditulis sebagai berikut:

$$X = X_{t-1} + X_t \tag{A}$$

Ket:

- X_{t-1} = Nilai kWh sebelumnya
- X_t = Nilai kWh saat ini
- X = Nilai kWh yang dicari

Sistem matematis 2.1 digunakan untuk mencari nilai *stationary* agar data *time series* yang digunakan dapat diubah ke suatu bentuk pola.

B. Clustering

Clustering merupakan suatu proses untuk mengelompokkan atau membagi objek yang sama ke dalam suatu *cluster*. *Clustering* sangat berguna untuk menemukan grup atau grup yang tidak dikenal dalam data. Ada berbagai macam metode yang dapat digunakan untuk melakukan *clustering* contohnya yaitu K-Means, K-Means++, DBSCAN, BIRCH, AHC (*Agglomerative Hierarchical Clustering*). Metode *clustering* terdiri dari 2 jenis yaitu *hierarchical clustering* dan *non-hierarchical clustering*. *Hierarchical clustering* adalah metode pengelompokan dengan menggabungkan dua *cluster* terdekat. Algoritma *clustering* ini akan berakhir ketika hanya ada satu *cluster* yang tersisa *Non*

Hierarchical Clustering melibatkan pembentukan *cluster* baru dengan menggabungkan atau memisahkan *cluster*. Teknik ini mengelompokkan data untuk memaksimalkan atau meminimalkan beberapa kriteria evaluasi [9].

C. Mini Batch K-Means

Algoritma pengelompokan Mini Batch K-Means adalah pengembangan dari algoritma K-means yang dapat digunakan sebagai pengganti algoritma K-means saat mengelompokkan pada kumpulan data besar. Algoritma ini bertujuan untuk memperbaiki kekurangan dan menyempurnakan algoritma K-Means, terkadang kinerjanya lebih baik daripada algoritma K-means standar saat bekerja pada kumpulan data besar, karena tidak mengulangi seluruh kumpulan data. Hal ini menciptakan kumpulan data acak untuk disimpan dalam memori, kemudian kumpulan data acak dikumpulkan pada setiap iterasi untuk memperbarui *cluster* dan ini diulang sampai konvergen[10]. Untuk menentukan probabilitas dari terpilihnya sebuah *centroid*, maka digunakan suatu metode yang dinamakan *weighting* dengan persamaan rumus sebagai berikut [11]:

$$K = \frac{D(i)}{\sum_{i \in D} D(i)} \tag{B}$$

Keterangan:

- K = Probabilitas nilai *centroid*
- $D(i)^2$ = Squared Distance
- $\sum_{i \in D} D(i)^2$ = Jumlah total nilai squared distance

Algoritma ini akan mengambil kumpulan dataset kecil yang dipilih secara acak untuk setiap iterasi. Langkah-langkah pengelompokkan data menggunakan algoritma Mini Batch K-Means adalah sebagai berikut:

1. Membagi set data menjadi beberapa *batch* secara acak.
2. Mengatur jumlah *cluster-cluster* awal.
3. Perbarui pusat cluster secara iteratif sampai pusat cluster tidak berubah lagi.

Adapun langkah-langkah pengelompokkan data menggunakan algoritma K-Means, yaitu:

1. Meletakkan titik K ke dalam ruang yang direpresentasikan oleh data yang dikelompokkan. Titik-titik ini akan mewakili *centroid* kelompok awal.
2. Menetapkan setiap objek ke grup yang memiliki pusat terdekat.
3. Hitung ulang posisi *centroid* K menggunakan metode *Euclidean*

Distance, setelah semua objek ditetapkan.

4. Ulangi point kedua dan ketiga hingga *centroid* tidak lagi bergerak.

Berikut ini merupakan persamaan untuk metode *Euclidean Distance* yang digunakan untuk menghitung ulang posisi *centroid* K.

$$D_{k_i} = \sqrt{(D_{k_i} - D_{k_i})^2 + (D_{k_i} - D_{k_i})^2} \quad (2.3)$$

Keterangan:

D_{k_i} = *Euclidean Distance*

i = Banyaknya objek

(D, D) = Koordinat objek

(D, D) = Koordinat *centroid*

D. Silhouette Coefficient

Silhouette Coefficient atau skor siluet adalah metode validasi konsistensi yang digunakan untuk menghitung kebaikan suatu teknik *clustering* pada pengelompokan data. Nilainya berkisar dari -1 hingga 1. Jika nilai *silhouette coefficient* mendekati nilai 1,

maka semakin baik pengelompokan data yang dilakukan didalam satu *cluster*, namun apabila nilainya mendekati -1 maka semakin buruk pengelompokan data pada suatu *cluster*. Metode ini akan menghitung koefisien siluet dari setiap titik yang mengukur seberapa mirip suatu titik dengan *clusternya* sendiri, dibandingkan dengan *cluster* lainnya dengan memberikan representasi grafis ringkas tentang seberapa baik setiap objek telah diklasifikasikan [12]. Adapun tahapan untuk menghitung *silhouette coefficient* adalah sebagai berikut:

1. Menghitung jarak rata-rata dari data ke-i ke semua data dalam satu cluster.
2. Hitung nilai dari rata-rata jarak data ke-i ke semua data yang ada di dalam cluster yang berbeda.
3. Hitung nilai *silhouette coefficient*.

$$S(i) = \frac{D(i) - D(i)}{m \{ D(i) - D(i) \}} \quad (D)$$

Keterangan:

$S(i)$ = *Silhouette index* data ke-i

$D(i)$ = Rata-rata jarak data ke-i dalam cluster index data ke-i

$S(i)$ = Nilai minimum dari rata-rata jarak data ke-i terhadap semua data dari cluster diluar dari satu cluster dengan data ke-i.

Nilai	Interpretasi Silhouette Coefficient
0,71 – 1,00	Struktur Kuat
0,51 – 0,70	Struktur Baik
0,26 – 0,50	Struktur Lemah
≤ 0,25	Struktur Buruk

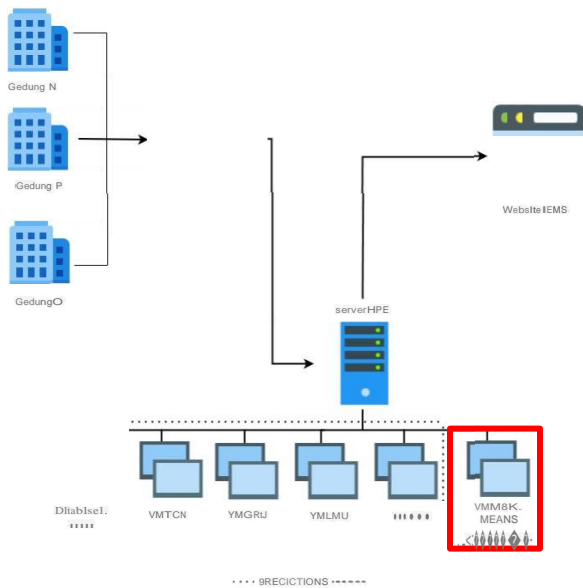
II. METODE

Sistem yang dirancang adalah system

pengelompokan data penggunaan energi listrik menggunakan algoritma Mini Batch K-Means. Hasil dari clustering yang dilakukan akan ditampilkan pada website IEMS.

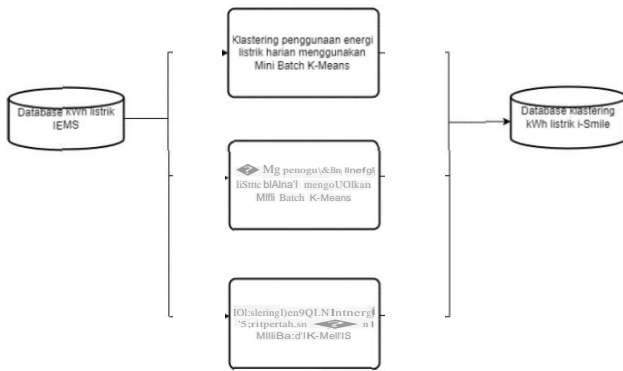
A. Desain Sistem

Pada gambaran umum sistem ini, data yang digunakan merupakan data penggunaan energi listrik di Fakultas Teknik Elektro Universitas Telkom. Data ini merupakan data yang diambil melalui database IEMS menggunakan API, selanjutnya dilakukan pengolahan data terlebih dahulu melalui program yang telah dibuat dan data disimpan ke dalam database laboratorium i-Smile setiap pukul 23.55 WIB. Data tersebut akan diolah kembali agar dapat menghasilkan hasil *clustering* dan hasilnya akan disimpan kembali ke dalam database laboratorium i-Smile dan akan ditampilkan pada website yang telah dirancang. *Website* ini berfungsi untuk menampilkan sejumlah informasi mengenai penggunaan energi listrik harian, bulanan dan pergedung di Gedung N, Gedung P dan Gedung O Fakultas Teknik Elektro.



Gambar 1 Gambaran Umum Sistem

Penelitian ini berfokus pada *clustering* penggunaan energi listrik menggunakan metode Mini Batch K-Means. *Clustering* dilakukan untuk mendapatkan hasil *clustering* perhari, perbulan dan pertahun menggunakan data yang diambil melalui database i-Smile, kemudian keluaran yang dihasilkan dikirim kembali ke database i-Smile.



Gambar 2 Gambaran Khusus Sistem

B. Sumber Data

1. Data Gedung N

Data yang digunakan adalah data listrik Gedung N Fakultas Teknik Elektro Universitas Telkom yang dimulai dari bulan September 2021 hingga bulan Mei 2022. Data yang diambil tersebut merupakan data penggunaan energi listrik yang di catat setiap 5 menit sekali. Kedua data tersebut diambil melalui server API Lab IEMS, dari API tersebut akan diambil data “Date” dan “Time” sebagai tanggal dan waktu serta “kWh” sebagai nilai dari beban listrik.

```

"message": "Berhasil aemuat data perangkat...",
"data": {
  "id": "137428",
  "id_device": "S...",
  "ve": "234.6...",
  "vb": "235.S...",
  "vc": "23s.1...",
  "vab": "e...",
  "vbc": "e...",
  "vcc": "e...",
  "ia": "1s.4...",
  "jb": "11.1...",
  "ic": "6.6...",
  "pa": "0.9...",
  "pb": "1...",
  "pc": "1...",
  "pt": "2.91...",
  "qa": "0...",
  "qb": "e...",
  "qc": "0...",
  "qt": "2.s1...",
  "ae": "4.18...",
  "sb": "2.43...",
  "sc": "1.19...",
  "st": "3.83...",
  "pfo": "e.22...",
  "pfb": "0.41...",
  "pfc": "0.84...",
  "freq": "s0...",
  "ep": "8737...",
  "eq": "9.83..."
},
"date": "2021-11-18",
"tice": "14:03:04"
    
```

Gambar 3 Data Raw JSON

2. Data dummy Gedung P dan Gedung O

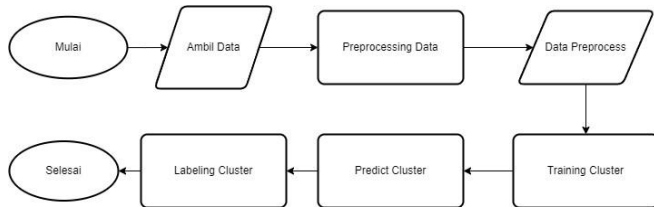
Dataset ini dibuat menggunakan bahasa pemrograman python yang akan selalu mengirimkan data random ke database, dan data ini memiliki kemiripan karakteristik seperti data Gedung N yang diambil melalui API untuk device 7 Gedung N yang sudah melalui proses *preprocessing*, data ini akan mengirimkan data yang sama ke Gedung P device 5 dan device 6 serta Gedung O.

The screenshot shows the phpMyAdmin interface for a database named 'i-smile'. A table is displayed with columns for 'date', 'time', and 'kWh'. The table contains multiple rows of data, including timestamps and numerical values representing electricity usage.

Gambar 4 Tampilan database virtual device

C. Rancangan Mini Batch K-Means

Dalam pembuatan sistem *clustering* penggunaan energi listrik ini diperlukan proses dari mengambil data dan membersihkan data yang diambil dari database untuk digunakan sebagai masukan dari model Mini Batch K-Means. Data yang sudah melalui proses *preprocessing*, akan digunakan untuk mencari nilai kenaikan penggunaan listrik.



Gambar 5 Alur Kerja Mini Batch K-Means Pada Sistem

Data yang sudah melalui proses *preprocessing* akan dilatih untuk menentukan pengelompokan data berdasarkan delta kWh. Selanjutnya dilakukan *predict cluster* untuk menentukan nilai *silhouette score* dari pengelompokan data yang dilakukan, data yang telah ditentukan *clusternya* kemudian dicari *range* nilai kWh dalam *cluster* yang dihasilkan untuk mengetahui data penggunaan listrik tersebut termasuk ke dalam penggunaan tinggi, rendah atau normal. Hasil akhir dari *clustering* ini akan ditampilkan pada website dan akan divisualisasikan menggunakan *doughnut chart* agar para pengguna dapat mengetahui kondisi pemakaian listrik pada saat itu.

IV. HASIL DAN PEMBAHASAN

Bagian ini membahas mengenai hasil pengujian pada pengelompokan data penggunaan energi listrik menggunakan algoritma Mini Batch K-Means Clustering. Berikut ini merupakan beberapa pengujian yang telah dilakukan:

1. Pengujian Pengelompokan Data Perbulan

Pengujian ini menggunakan data Gedung N dari Gedung N Fakultas Teknik Elektro dan data dummy Gedung P dan Gedung O, pengujian ini berdasarkan hari dalam satu bulan yaitu 30 hari. Hasil *clustering* akan divisualisasikan pada website dengan ketentuan yaitu:

1. Warna biru memiliki nilai kenaikan kWh yang tinggi
2. Warna hijau memiliki nilai kenaikan kWh yang normal
3. Warna kuning memiliki nilai kenaikan kWh yang rendah

Pada pengujian ini didapatkan hasil *silhouette score* sebesar 0,599 untuk data Gedung N dan 0,452 untuk data dummy untuk Gedung P dan Gedung O. Hasil dari *silhouette score* data Gedung N termasuk ke dalam struktur baik, sedangkan hasil *silhouette score* Data dummy Gedung P dan Gedung O.

```

silhouette_score(dk
0.5755337900397074
  
```

Gambar 6 Silhouette Score Data Gedung N Perbulan

```

silhouette_score(dk
0.5045078350031963
  
```

Gambar 7 Silhouette Score Data dummy Gedung P Dan Gedung O Perbulan

Setelah diketahui nilai *silhouette score* dari kedua data tersebut, maka perlu dilakukan pencarian *range* delta kWh tiap *cluster* agar dapat diketahui mana yang tergolong penggunaan rendah, normal dan tinggi. Hal tersebut dapat diamati melalui nilai minimal dan maksimal delta kWh dari setiap *cluster* yang terbentuk. Nilai-nilai yang terbentuk atau nilai delta kWh berasal dari nilai kWh satu jam sebelumnya dikurangi dengan nilai kWh sekarang. Adapun nilai-nilai tersebut untuk Gedung N adalah sebagai berikut:

• Clustering Perbulan

Cluster 0 (normal) : 74,4 – 118,5

Cluster 1 (rendah) : 20,0 – 71,39

Cluster 2 (tinggi) : 119,0 – 195,0

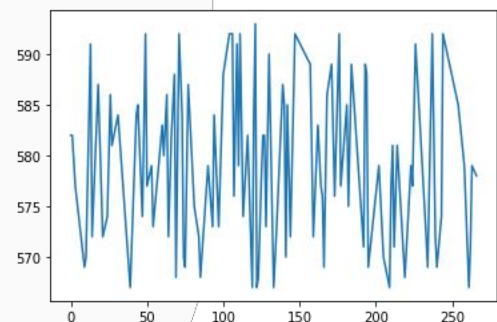
Sedangkan untuk Gedung P dan Gedung O adalah sebagai berikut:

• Clustering Perbulan

Cluster 0 (normal) : 567 – 593

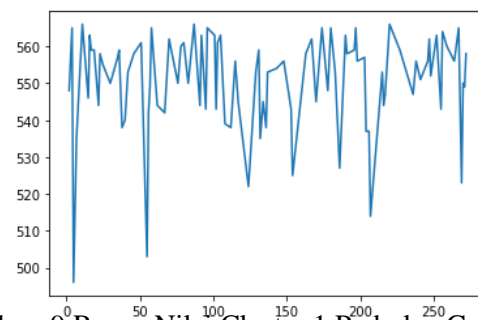
Cluster 1 (rendah) : 496 – 566

Cluster 2 (tinggi) : 594 – 640



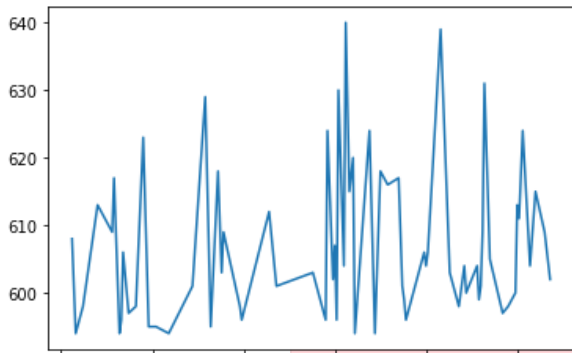
Gambar 8 Range Nilai Cluster 0 Perbulan Gedung P dan Gedung O

Pada gambar 8 terdapat grafik dari hasil *cluster 0* perbulan pada gedung P dan gedung O yang dimulai dari range antara 567 sampai dengan 593 dan termasuk ke dalam kategori normal.



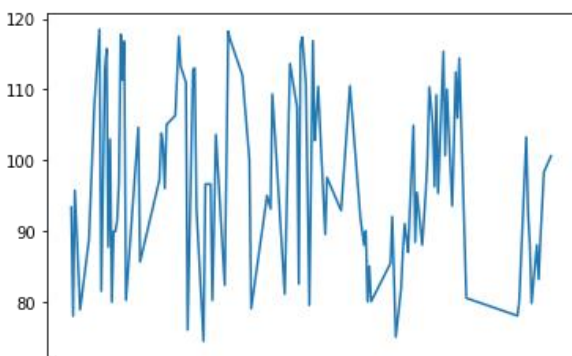
Gambar 9 Range Nilai Cluster 1 Perbulan Gedung P dan Gedung O

Pada gambar 9 terdapat grafik dari hasil *cluster 1* perbulan pada gedung P dan gedung O yang dimulai dari range antara 496 sampai dengan 566 dan termasuk ke dalam kategori rendah.



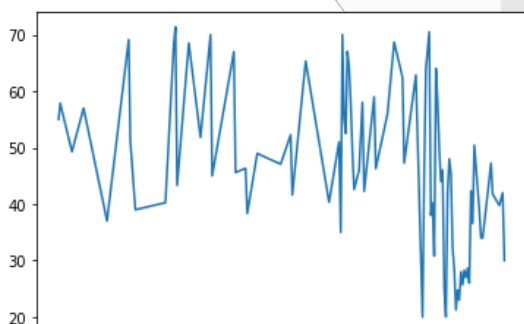
Gambar 10 Range Nilai Cluster 2 Perbulan Gedung P dan Gedung O

Pada gambar 10 terdapat grafik dari hasil *cluster 2* perbulan pada gedung P dan gedung O yang dimulai dari range antara 594 sampai dengan 640 dan termasuk ke dalam kategori tinggi.



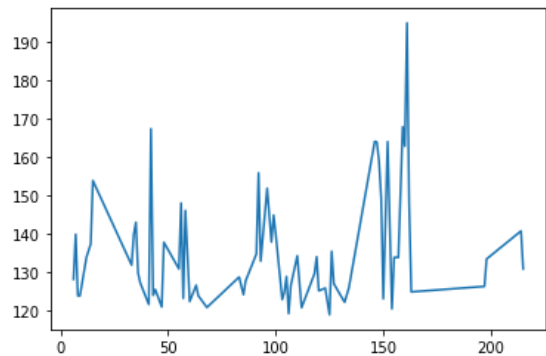
Gambar 11 Range Nilai Cluster 0 Perbulan Gedung N

Pada gambar 11 terdapat grafik dari hasil *cluster 0* perbulan pada gedung N yang dimulai dari range antara 74,4 sampai dengan 118,5 dan termasuk ke dalam kategori normal.



Gambar 12 Range Nilai Cluster 1 Perbulan Gedung N

Pada gambar 12 terdapat grafik dari hasil *cluster 1* perbulan pada gedung N yang dimulai dari range antara 20,0 sampai dengan 71,39 dan termasuk ke dalam kategori rendah.



Gambar 13 Range Nilai Cluster 2 Perbulan Gedung N Pada gambar 13 terdapat grafik dari hasil *cluster 2* perbulan pada gedung N yang dimulai dari range antara 119,0 sampai dengan 195,0 dan termasuk ke dalam kategori tinggi.

2. Pengujian Pengelompokan Data Perhari

Pengujian ini juga menggunakan data Gedung N dan data dummy Gedung P dan Gedung O, pengujian ini berdasarkan jam dalam satu hari yaitu 24 jam maka pengujian ini menggunakan data sebanyak 24 data dalam satu hari. Hasil pengujian pada pengelompokan data perhari akan di visualisasikan pada website. Pada pengujian ini didapatkan hasil *silhouette score* sebesar 0,62 untuk data Gedung N dan 0,55 untuk data data dummy Gedung P dan Gedung O. Kedua hasil ini termasuk ke dalam struktur baik.

```
silhouette_score(dk,
```

Gambar 14 Silhouette Score Data Gedung N Perhari

```
silhouette_score(dk,
```

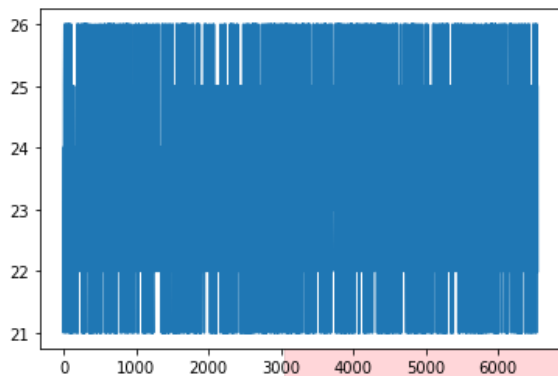
Gambar 15 Silhouette Score Data Dummy Gedung P dan Gedung O Perhari

Setelah diketahui nilai *silhouette score* dari kedua data tersebut, maka perlu dilakukan pencarian *range delta kwh* tiap *cluster* agar dapat diketahui mana yang tergolong penggunaan rendah, normal dan tinggi. Nilai-nilai yang terbentuk atau nilai delta kWh berasal dari nilai kWh satu jam sebelumnya dikurangi dengan nilai kWh sekarang. Hal tersebut dapat diamati melalui nilai minimal dan maksimal delta kWh dari setiap *cluster* yang terbentuk. Adapun nilai-nilai tersebut untuk Gedung N adalah sebagai berikut:

- *Clustering Perhari*
Cluster 0 (normal) : 3,62 – 8,12
Cluster 1 (rendah) : 0,0– 3,54
Cluster 2 (tinggi) : 8,16 – 29,0

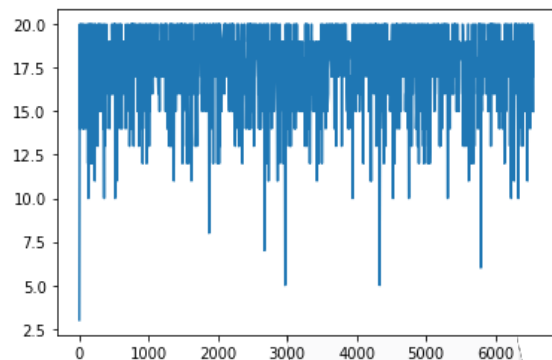
Sedangkan untuk Gedung P dan Gedung O adalah sebagai berikut:

- *Clustering* Perhari
Cluster 0 (normal) : 21 – 26
Cluster 1 (rendah) : 3 – 20
Cluster 2 (tinggi) : 27 – 41



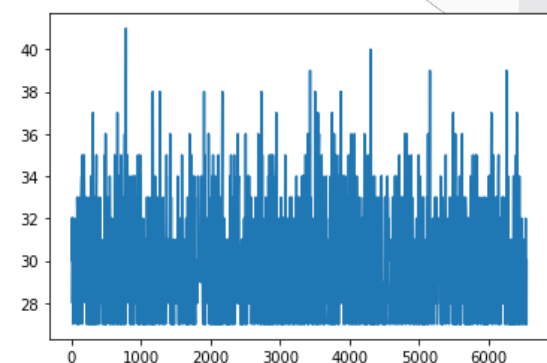
Gambar 16 Range Nilai Cluster 0 Perhari Gedung P dan Gedung O

Pada gambar 16 terdapat grafik dari hasil *cluster 0* perhari pada gedung P dan gedung O yang dimulai dari range antara 21 sampai dengan 26 dan termasuk ke dalam kategori normal.



Gambar 17 Range Nilai Cluster 1 Perhari Gedung P dan Gedung O

Pada gambar 17 terdapat grafik dari hasil kluster 1 perhari pada gedung P dan gedung O yang dimulai dari range antara 3 sampai dengan 20 dan termasuk ke dalam kategori rendah.



Gambar 18 Range Nilai Cluster 2 Perhari Gedung P dan Gedung O

Pada gambar 18 terdapat grafik dari hasil kluster 2 perhari pada gedung P dan gedung O yang dimulai dari range antara 27 sampai dengan 41 dan termasuk ke dalam kategori tinggi.

3. Pengujian Pengelompokkan Data Pertahun

Pengujian ini dilakukan untuk mengetahui nilai *silhouette score* dari bulan September 2021 sampai dengan Mei 2022. Berdasarkan hasil pengujian yang dilakukan, didapatkan hasil *silhouette score* sebesar 0,527 untuk data Gedung N dan 0,731 untuk data dummy Gedung P dan Gedung O. Kedua hasil yang didapat termasuk ke dalam kategori struktur kuat.

```
silhouette_score(dk_
0.7316454766637361
```

Gambar 19 Silhouette Score Data Gedung N Pertahun

```
silhouette_score(dk_
0.7204653851611652
```

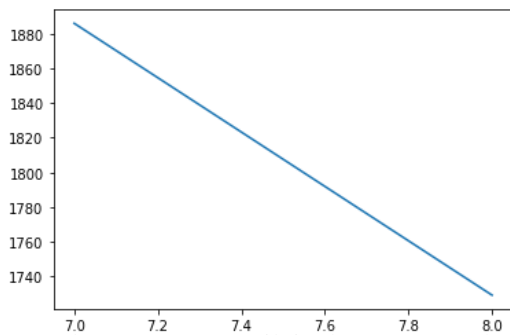
Gambar 20 Silhouette Score Data dummy Gedung P dan Gedung O Pertahun

Setelah diketahui nilai *silhouette score* dari kedua data tersebut, maka perlu dilakukan pencarian *range delta kwh* tiap *cluster* agar dapat diketahui mana yang tergolong penggunaan rendah, normal dan tinggi. Nilai-nilai yang terbentuk atau nilai delta kWh berasal dari nilai kWh satu jam sebelumnya dikurangi dengan nilai kWh sekarang. Hal tersebut dapat diamati melalui nilai minimal dan maksimal delta kWh dari setiap *cluster* yang terbentuk. Adapun nilai-nilai tersebut untuk Gedung N adalah sebagai berikut:

- *Clustering* Pertahun
Cluster 0 (normal) : 1729,32–1885,95
Cluster 1 (rendah) : 2624,45–2999,0
Cluster 2 (tinggi) : 3345,69–3458,5

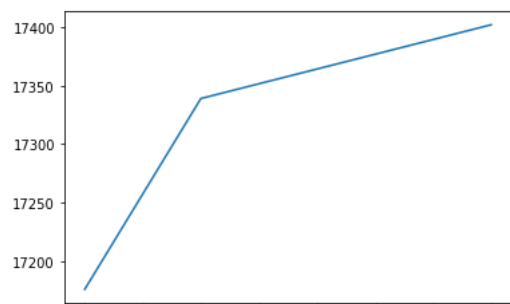
Sedangkan untuk Gedung P dan Gedung O adalah sebagai berikut:

- *Clustering* Pertahun
Cluster 0 (normal) : 17176–17402
Cluster 1 (rendah) : 17770–17899
Cluster 2 (tinggi) : 16322



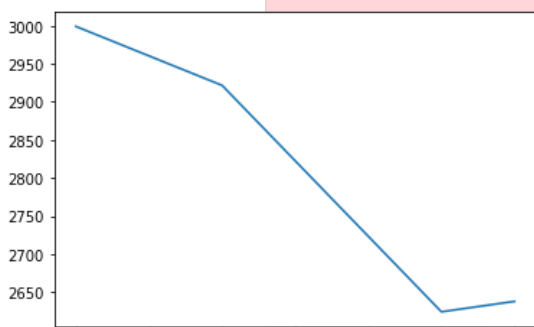
Gambar 21 Range Nilai Cluster 0 Perbulan Gedung N

Pada gambar 21 terdapat grafik dari hasil kluster 0 pertahun pada gedung N yang dimulai dari range antara 1729,32 sampai dengan 1885,95 dan termasuk ke dalam kategori normal.



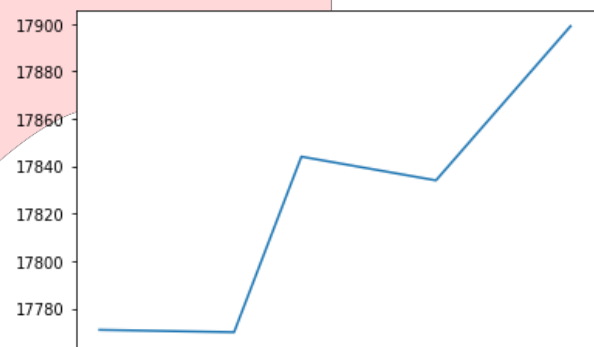
Gambar 24 Range Nilai Cluster 0 Perbulan Gedung P dan Gedung O

Pada gambar 24 terdapat grafik dari hasil kluster 0 pertahun pada gedung Gedung P dan Gedung O yang dimulai dari range antara 1716 sampai dengan 17402 dan termasuk ke dalam kategori normal.



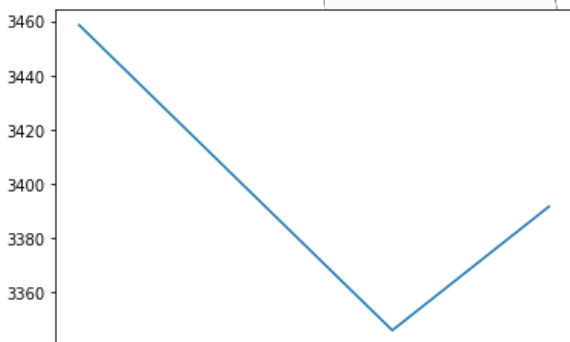
Gambar 22 Range Nilai Cluster 1 Perbulan Gedung N

Pada gambar 22 terdapat grafik dari hasil kluster 1 pertahun pada gedung N yang dimulai dari range antara 2624,45 sampai dengan 2999,0 dan termasuk ke dalam kategori rendah.



Gambar 25 Range Nilai Cluster 1 Perbulan Gedung P dan Gedung O

Pada gambar 25 terdapat grafik dari hasil kluster 1 pertahun pada gedung Gedung P dan Gedung O yang dimulai dari range antara 2624,45 sampai dengan 2999,0 dan termasuk ke dalam kategori rendah. Untuk tampilan grafik range nilai cluster 2 pertahun pada Gedung P dan Gedung O tidak dapat ditampilkan dikarenakan hanya ada satu data yang termasuk ke dalam cluster 2.



Gambar 23 Range Nilai Cluster 2 Perbulan Gedung N

Pada gambar 23 terdapat grafik dari hasil kluster 2 pertahun pada gedung N yang dimulai dari range antara 3345,69 sampai dengan 3458,5 dan termasuk ke dalam kategori tinggi.

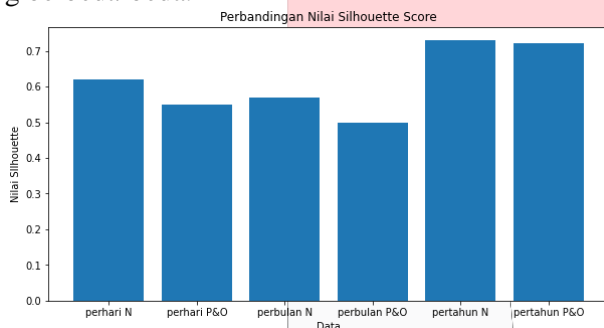
4. Pengujian Silhouette Coefficient

Pengujian ini dilakukan untuk mengetahui seberapa baik atau burukkah pengelompokan data yang dilakukan, dengan pengujian tersebut maka didapatkan hasil sebagai berikut:

No	Hasil Pengelompokan	Silhouette Coefficient
1	Perbulan Data Gedung N	0,57
2	Perbulan Data dummy Gedung P dan Gedung O	0,50
3	Perhari Data Gedung N	0,62

No	Hasil Pengelompokkan	Silhouette Coefficient
4	Perhari Data dummy Gedung P dan Gedung O	0,55
5	Pertahun Data Gedung N	0,73
6	Pertahun Data dummy Gedung P dan Gedung O	0,72

Dari pengujian yang dilakukan maka dapat disimpulkan bahwa antara data Gedung N dan Data dummy Gedung P dan Gedung O yang digunakan memiliki kemiripan yang tinggi, karena dari hasil *silhouette score* berdasarkan pengelompokkannya masing-masing semuanya termasuk ke dalam struktur yang berbeda-beda.



Gambar 26 Perbandingan Nilai Silhouette Score

V. KESIMPULAN

Berdasarkan analisa dari perancangan dan pengujian yang dilakukan, maka diambil kesimpulan sebagai berikut:

1. Website dapat menampilkan hasil *clustering* dengan baik untuk pengguna.
2. Hasil dari pengujian yang telah dilakukan menunjukkan bahwa nilai *silhouette score* berdasarkan *clustering* perhari menggunakan data Gedung N bernilai 0,62 dan hasil *clustering* perbulan data Gedung N sebesar 0,57, seluruh hasil tersebut termasuk ke dalam struktur baik dan hasil *clustering* pertahun data Gedung N sebesar 0,73 termasuk ke dalam struktur kuat. Hasil *clustering* menggunakan data dummy Gedung P dan Gedung O adalah sebesar 0,55 untuk perhari yang termasuk ke dalam struktur baik, perbulan sebesar 0,50 termasuk ke dalam struktur lemah dan pertahun sebesar 0,72 termasuk ke dalam struktur kuat.

REFERENSI

- [1] Ressay Aryani, Muhammad Nasrun, Casi Setianingsih, and Muhammad Ary Murti, "Clustering Data In Power Management Sistem Using K-Means Algorithm" *Proceedings, 2019 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob): 5-7 November 2019, Bali, Indonesia*. IEEE, 2019.
- [2] Liping Zhang, Song Deng, and Shiyue Li, "Analysis of Power Consumer Behavior Based on the Complementation of K-Means and DBSCAN," *Conference on Energy Internet and Energy Sistem Integration (E12)*, 2017.
- [3] A. Shokrollahi and B. Mazloom-Nezhad Maybodi, "An energy-efficient clustering algorithm using fuzzy c-means and genetic fuzzy sistem for wireless sensor network," *Journal of Circuits, Systems and Computers*, vol. 26, no. 1, Jan. 2017, doi: 10.1142/S0218126617500049.
- [4] Y. Amri, A. L. Fadhillah, Fatmawati, N. Setiani, and S. Rani, "Analysis Clustering of Electricity Usage Profile Using K-Means Algorithm," in *IOP Conference Series: Materials Science and Engineering*, Jan. 2016, vol. 105, no. 1. doi: 10.1088/1757-899X/105/1/012020.
- [5] A. M. Riad, I. Elhenawy, A. Hassan, and N. Awadallah, "Visualize Network Anomaly Detection by Using K-Means Clustering Algorithm," *International journal of Computer Networks & Communications*, vol. 5, no. 5, pp. 195–208, Sep. 2013, doi: 10.5121/ijcnc.2013.5514.
- [6] N. R. Hikmiyah, R. R. A. Siregar, B. Prayitno, D. T. Kusuma, and N. G. Pahiyanti, "Metode Fuzzy Subtractive Clustering Dalam Pengelompokkan Penggunaan Energi Listrik Rumah Tangga," *Petir*, vol. 14, no. 2, pp. 269–279, Sep. 2021, doi: 10.33322/petir.v14i2.1448.
- [7] Marco Peixero, "The Complete Guide to Time Series Analysis and Forecasting," *towards data science*, Aug. 07, 2019. <https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775> (accessed Oct. 28, 2021).
- [8] Agnes Lisnawati, "Model Exponential Smoothing Holt-Winter dan Model SARIMA untuk peramalan tingkat hunian hotel di provinsi DIY," *UNY Journal*, 2019.
- [9] GeeksforGeeks, "Difference between Hierarchical and Non Hierarchical Clustering." <https://www.geeksforgeeks.org/difference-between-hierarchical-and-non-hierarchical->

clustering/#:~:text=Non%20Hierarchical%20Clustering%3A%201.%20Hierarchical%20Clustering%20involves%20creating,is%20considered%20less%20reliable%20than%20Non%20Hierarchical%20Clustering. (accessed Dec. 03, 2021).

[10] GeeksforGeeks, "ML | Mini Batch K-means clustering algorithm," *GeeksforGeeks*, 2019. <https://www.geeksforgeeks.org/ml-mini-batch-k-means-clustering-algorithm/> (accessed Dec. 02, 2021).

[11] Siti Rofiqoh Fitriyani and Hendri Murfi, "The K-Means with Mini Batch Algorithm for Topics Detection on Online News," *2016 Fourth International Conference on Information and Communication Technologies (ICoICT)*, 2016.

[12] Satyam Kumar, "Silhouette Method — Better than Elbow Method to find Optimal

Clusters," *Towards Data Science*, Sep. 19, 2020. <https://towardsdatascience.com/silhouette-method-better-than-elbow-method-to-find-optimal-clusters-378d62ff6891> (accessed Dec. 03, 2021).

[13] S. Darma and S. Sistem, "Studi Sistem Peneraan KWH Meter," 2019.

[14] Unzhil Latif Jayyid, "Analisis Penggunaan KWH Meter Pascabayar Dan KWH Meter Prabayar 1 Fasa Di PT. PLN (PERSERO)," Medan, 2016.

[15] Bobby, "Daftar Tarif Listrik Terbaru 2021 Kementerian ESDM [Harga kWh]," *lifepal*, Jun. 07, 2021. <https://lifepal.co.id/media/daftar-tarif-listrik-terbaru/> (accessed Nov. 18, 2021).

