

Implementasi Algoritma K-Means Clustering Penggunaan Sumber Daya BBM Pada Genset Di PT Telkom Arnet Medan

Monika Mastiur Panjaitan¹, Asaziduhu Gea², Surianto Sitepu³, Naikson F Saragih⁴, Indra M Sarkis⁵

^{1,2,3}Fakultas Ilmu Komputer, Universitas Methodist Indonesia

Info Artikel

Histori Artikel:

Received, Jul 20, 2022

Revised, Jul 30, 2020

Accepted, Aug 15, 2020

Keywords:

*K-Means,
Clustering,
Davies Bouldin,
Penggunaan BBM*

ABSTRAK

Penelitian ini mendeskripsikan dan mengkaji mengenai permasalahan, pertama mengenai bagaimana menerapkan data minning dalam menganalisis data pada penggunaan jam operasi genset tertinggi dan terendah. Kedua, bagaimana menerapkan algoritma K-Means Clustering dalam pengolahan data dan menghasilkan informasi terhadap jam operasi dan pemakaian bbm pada genset. Penelitian ini menggunakan K-Means Clustering bertujuan untuk memberikan pengetahuan baru mengenai informasi genset menggunakan algoritma K-Means Clustering dan mengetahui penggunaan jam operasi genset tertinggi dan terendah dari setiap STO yang terdapat pada Telkom Arnet Medan. Berdasarkan data yang didapat dari hasil peng-clusteran, hasil yang di dapat dalam penelitian ini dapat disimpulkan bahwa : 1) pengujian dilakukan dalam 2 kali iterasi clustering pada data genset; 2) data diolah menggunakan rapidminer untuk ditentukan centroid dalam 2 cluster yaitu cluster untuk penggunaan jam operasi genset tertinggi (C1) dan cluster untuk penggunaan jam operasi terendah (C2); 3) Dari data 23 lokasi genset didapatkan 1 STO cluster penggunaan jam operasi genset tertinggi yaitu STO Tembung japati, 22 STO cluster penggunaan jam operasi genset terendah; 4) Nilai Davies Bouldin yang dihasilkan dari proses validitas cluster sebesar 0.036

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Penulis Koresponden:

Monika Mastiur Panjaitan,
Faculty of Computer Science,
Universitas Methodist Indonesia, Medan,
Jl. Hang Tua No.8, Medan - Sumatera Utara.
Email: monikampanjaitan98@gmail.com

1. PENDAHULUAN

Divisi Operational and Maintenance (CME & Data Center) merupakan salah satu divisi pendukung di lingkungan PT Telkom Arnet Medan yang bertanggungjawab penuh atas seluruh perangkat yang ada di 23 STO (Sentral telepon Otomat). Banyaknya perangkat yang harus ditangani, sehingga perusahaan sangat dituntut adanya pengelolaan dan pengawasan yang baik terhadap perangkat dimulai dari perencanaan, pemasangan, pengujian, pengoperasian, sampai pemeliharaan, sehingga perangkat yang digunakan perusahaan tersebut aman dan efisien. Salah satu perangkat yang sangat penting yang ada di perusahaan adalah genset. Genset di gunakan sebagai pengganti tenaga listrik untuk menunjang segala aktivitas yang akan dilakukan perusahaan, genset pada perusahaan ini ada di setiap lokasi STO. Pemakaian bbm genset pada setiap STO berbeda – beda tergantung jam operasi dari genset tersebut.

Penambahan bbm pada genset dilakukan 3 hingga 6 sekali, banyaknya data yang harus dilihat setiap bulannya sehingga membuat pihak perusahaan lama dalam melakukan pengelompokkan data. Saat ini perusahaan menginginkan penambahan bbm di hitung berdasarkan jam operasi yang paling banyak di lakukan genset pada setiap STO. Hal tersebut dilakukan agar menghindari beberapa masalah yaitu meminimalisir terbuangnya solar apabila tiba – tiba pipa tangki genset bocor, bbm

solar yang terlalu lama di simpan di tangki dapat mudah rusak dan bisa menjadi lumpur yang dapat merugikan perusahaan. Untuk dapat melakukan penambahan BBM pada tangki genset yang lebih efektif dan efisien, maka dalam penelitian ini dilakukan dengan cara mengelola data – data yang telah didapatkan dari PT Telkom Arnet Medan mengenai genset seperti lokasi STO, pemakaian jam operasi dari genset, pemakaian solar perliternya. Data-data yang telah didapatkan kemudian diolah untuk mengetahui pola dari data-data tersebut sehingga peneliti dapat mengambil informasi-informasi yang tersembunyi dari data-data tersebut. Metode pengolahan data seperti ini sering disebut sebagai data *mining*.

Pada penelitian ini analisa data *mining* dilakukan dengan menggunakan metode *K-Means clustering*. Dengan menggunakan metode ini, data - data yang telah didapatkan dapat dikelompokkan kedalam beberapa cluster berdasarkan kemiripan dari data-data tersebut, sehingga data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster dan yang memiliki karakteristik yang berbeda dikelompokkan dalam *cluster* yang lain yang memiliki karakteristik yang sama. Penelitian ini diperkuat oleh penelitian Hajar et al., (2020) dengan judul “Penerapan K-Means Clustering Pada Ekspor Minyak Kelapa Sawit Menurut Negara Tujuan”. Dalam penelitian tersebut hasil analisis yang bertujuan untuk mengetahui variabel apa saja yang menjadi pengelompokkan cluster tingkat ekspor tinggi dan cluster tingkat ekspor rendah, dan diketahui Centroid data untuk cluster tingkat ekspor tinggi 5290,90 dan Centroid data untuk Centroid data untuk cluster tingkat ekspor rendah 12,70. Sehingga diperoleh penilaian berdasarkan indeks ekspor minyak kelapa sawit dengan 1 negara cluster tingkat ekspor tinggi yakni India dan 9 negara cluster tingkat ekspor rendah yakni Tiongkok, Singapura, Malaysia, Pakistan, Bangladesh, Sri Lanka, Mesir, Belanda, dan Jerman. Hal ini dapat menjadi masukan kepada pemerintah, negara yang menjadi prioritas tertinggi pada kegiatan ekspor minyak kelapa sawit berdasarkan klaster yang telah dilakukan.

Pada metode *k-means clustering*, proses pengelompokkan data dilakukan melalui sebuah proses perulangan dimana perulangan akan dihentikan pada saat posisi cluster awal sudah sama dengan posisi cluster pada saat perulangan terakhir dilakukan. Pada penelitian ini dilakukan pengelompokkan data genset pada setiap STO berdasarkan penggunaan jam operasi genset tertinggi dan terendah.

2. METODOLOGI PENELITIAN

2.1 Data Mining

Data *Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari database yang besar[1].

Data *Mining* merupakan suatu rangkaian proses untuk menggali nilai tambah dari sekumpulan data yang berupa pengetahuan yang selama ini tidak diketahui secara manual. Data *Mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan basis data. Beberapa teknik yang sering disebut-sebut dalam literatur Data *Mining* antara lain : *clustering*, *classification*, *association rule mining*, *neural network*, dan *genetic algorithm*.

2.2 Clustering

Salah satu teknik yang dikenal dalam data *Mining* yaitu *clustering*. Pengertian clustering keilmuan dalam data *Mining* adalah pengelompokan sejumlah data atau objek ke dalam *cluster* (*group*) sehingga setiap dalam *cluster* tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek dalam *cluster* yang lainnya.

Metode *clustering* yang paling banyak digunakan ialah metode *K-Means clustering*. Kelemahan utama dari metode ini adalah hasil yang sensitif terhadap pemilihan pusat cluster awal dan perhitungan solusi lokal untuk mencapai kondisi optimal. Analisis *cluster* merupakan teknik multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis *cluster* mengklasifikasi objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam *cluster* yang sama.

2.3 Algoritma K-Means

K-means merupakan salah satu metode yang sering digunakan untuk *clustering*. Algoritma *K-Means* merupakan salah satu metode *clustering* non hirarki yang memiliki waktu komputasi yang relatif cepat. Algoritma *K-Means* ini dimulai dengan menentukan terlebih dahulu berapa cluster yang ingin dibuat, selanjutnya menentukan nilai awal untuk masing-masing *cluster*, lalu menghitung jarak dari masing-masing data dengan nilai awal yang telah ditentukan. Setelah jarak dari tiap data dihitung, data dimasukkan dalam cluster terdekat. Selanjutnya menghitung rata-rata dari tiap *cluster*, rata-rata tersebut nantinya akan digunakan sebagai nilai awal yang baru dalam menghitung jarak dari masing-masing data.

Langkah-langkah melakukan *clustering* dengan metode *K-Means* adalah sebagai berikut :

1. Pilih jumlah *cluster* k .
2. Inisialisasi k pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat *cluster* diberi nilai awal dengan angka-angka random.
3. Alokasikan semua data/ objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$d(i,j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \dots\dots\dots(1)$$

dimana :

$d(i,j)$ = Jarak data ke i ke pusat *cluster* j

x_{ki} = Data ke i pada atribut data ke k

x_{kj} = Titik pusat ke j pada atribut ke k

4. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari *cluster* tersebut. Jadi rata-rata (*mean*) bukan satu-satunya ukuran yang bisa dipakai. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 (tiga) sampai pusat *cluster* tidak berubah lagi.

3. HASIL DAN PEMBAHASAN

3.1 Penerapan Algoritma K-means Clustering

Dalam penelitian ini, penulis menggunakan data gensek perangkat milik PT Telkom Arnet Medan. Data yang digunakan adalah data gensek yang terdapat pada 23 STO dan akan di olah menggunakan algoritma *K-Means*. Data yang digunakan dapat dilihat pada Gambar 1 berikut:

NO	LOKASI	PEMAKAIAN (JAM)	PEMAKAIAN (LTR)
1	Binjai (BJIA)	0.48	20
2	Cinta Damai (CTDA)	0.20	11
3	Delitua (DLTA)	0.20	1
4	Galang (GLGA)	2.05	4
5	Kuala (KULA)	0.40	4
6	Kuala Putri (KPTI)	4.41	61
7	Lubuk Pakam (LBPA)	0.47	15
8	Medan Centrum (MDNC)	0.20	63
9	Padang Bulan (PDBA)	0.20	9
10	Pangkalan Brandan (PBDA)	0.20	3
11	Pangkalan Susu (PSSA)	2.15	14
12	Perbaungan (PERA)	2.00	12
13	Percut (PRCA)	0.20	1
14	Pulo Brayan (PUBA)	0.20	21
15	Simpang Limun (SPLA)	2.05	58
16	Stabat (STBA)	7.43	123
17	Sukaramai (SKIA)	0.20	4
18	Tanjung Pura (TJPA)	1.28	9
19	Tanjung Morawa (TJRA)	4.00	76
20	Tanjung Mulia (TJMA)	2.20	44
21	Tembung (TMUA)	0.20	6
22	Tembung Japati (TBGB)	0.20	18
23	Tuntungan (TTGA)	1.20	11

Gambar 1 Data yang Akan Digunakan

3.2 Transformasi Data

Agar data diatas dapat diolah, maka data tersebut di transformasi terlebih dahulu agar memudahkan perhitungan menggunakan metode *k-means clustering* dapat dilihat pada tabel 1.

Tabel 1 Transformasi Pemakaian (Jam)

Pemakaian (Jam)	Transformasi
1 – 10	1
11-20	2
21 -30	3

Tabel 2 Transformasi (Ltr)

Pemakaian (Ltr)	Transformasi
10 – 500	1
600 – 1.000	2
1.100 – 1.500	3
1.600 – 2.000	4

3.3 Pengolahan Data

Setelah transformasi data, kemudian data yang ada pada tabel 1 akan di ubah sesuai ke dalam bentuk transformasi data di atas dan di olah ke dalam algoritma *K-Means*. Data dapat di lihat pada tabel 3.

Tabel 3 Data yang Sudah di Transformasi

No	Lokasi	Pemakaian (Jam)	Pemakaian (Ltr)
1	Binjai	1	1
2	Cinta Damai	1	1
3	Delitua	2	1
4	Galang	2	1
5	Kuala	2	1
6	Kuala Putri	2	1
7	Lubuk Pakam	2	1
8	Medan Centrum	1	1
9	Padang Bulan	1	1
10	Pangkalan Brandan	1	1
11	Pangkalan Susu	3	1
12	Perbaungan	2	1
13	Percut	1	1
14	Pulo Brayan	1	1
15	Simpang Limun	1	1
16	Stabat	2	1
17	Sukaramai	1	1
18	Tanjung Pura	3	1
19	Tanjung Morawa	1	1
20	Tanjung Mulia	2	1
21	Tembung	1	1
22	Tembung Japati	3	6
23	Tuntungan	1	1

Untuk dapat melakukan pengelompokan data-data tersebut menjadi beberapa *cluster* perlu dilakukan beberapa langkah, yaitu:

1. Menentukan Jumlah *cluster*, jumlah *cluster* merupakan jumlah kelompok yang akan dihasilkan. Dalam penelitian ini jumlah *cluster* yang akan digunakan adalah sebanyak 2 (dua) *cluster*.

2. Menentukan *centroid* awal, *centroid* awal diperoleh secara acak. *Centroid* awal merupakan titik pusat *cluster* pertama. Titik pusat awal setiap *cluster* dari penelitian ini dapat dilihat pada tabel 4.

Tabel 4 Titik Pusat Awal Setiap Cluster

	Lokasi	Pemakaian (Jam)	Pemakaian (Ltr)
C1	Percut	1	1
C2	Tembung Japati	3	6

3. Tempatkan setiap data pada *cluster*. Untuk mengetahui *cluster* mana yang paling dekat dengan data, maka perlu dihitung jarak setiap data dengan titik pusat setiap *cluster*. Sebagai contoh, akan dihitung jarak dari data genset pertama ke pusat *cluster* pertama:

Jarak data dengan *Centroid* C1 adalah

$$C1 = (1 \ 1)$$

$$1. D(x_1, c_1) = \sqrt{(a_1 - c_{1a})^2 + (b_1 - c_{1a})^2} = \sqrt{(1 - 1)^2 + (1 - 1)^2} = 0$$

$$2. D(x_2, c_1) = \sqrt{(a_2 - c_{1a})^2 + (b_2 - c_{1a})^2} = \sqrt{(1 - 1)^2 + (1 - 1)^2} = 0$$

$$3. D(x_3, c_1) = \sqrt{(a_3 - c_{1a})^2 + (b_3 - c_{1a})^2} = \sqrt{(2 - 1)^2 + (1 - 1)^2} = 1$$

Jarak data dengan *Centroid* C2 adalah

$$C2 = (3 \ 6)$$

$$1. D(x_1, c_2) = \sqrt{(a_1 - c_{2a})^2 + (b_1 - c_{2a})^2} = \sqrt{(1 - 3)^2 + (1 - 6)^2} = 5.38$$

$$2. D(x_2, c_2) = \sqrt{(a_2 - c_{2a})^2 + (b_2 - c_{2a})^2} = \sqrt{(1 - 3)^2 + (1 - 6)^2} = 5.38$$

$$3. D(x_3, c_2) = \sqrt{(a_3 - c_{2a})^2 + (b_3 - c_{2a})^2} = \sqrt{(2 - 3)^2 + (1 - 6)^2} = 5.09$$

Dari perhitungan diatas kemudian dikelompokkan berdasarkan tabel dan hasil *clustering* nya dengan menentukan nilai terendah dari 2 (dua) *centroid* yang dipilih. Hasil analisa *cluster iterasi* 1 dapat dilihat pada tabel 5.

Tabel 5 Hasil Analisa Cluster Iterasi 1

No	Lokasi	Jarak ke		Jarak Terdekat
		C1	C2	
1	Binjai	0	5.38	C1
2	Cinta Damai	0	5.38	C1
3	Delitua	1	5.09	C1
4	Galang	1	5.09	C1
5	Kuala	1	5.09	C1
6	Kuala Putri	1	5.09	C1
7	Lubuk Pakam	1	5.09	C1
8	Medan Centrum	0	5.38	C1
9	Padang Bulan	0	5.38	C1
10	Pangkalan Brandan	0	5.38	C1
11	Pangkalan Susu	2	5	C1
12	Perbaungan	1	5.09	C1
13	Percut	0	5.38	C1
14	Pulo Brayon	0	5.38	C1
15	Simpang Limun	0	5.38	C1
16	Stabat	1	5.09	C1
17	Sukaramai	0	5.38	C1
18	Tanjung Pura	2	5	C1
19	Tanjung Morawa	0	5.38	C1
20	Tanjung Mulia	1	5.09	C1
21	Tembung	0	5.38	C1
22	Tembung Japati	5.38	0	C2
23	Tuntungan	0	5.38	C1

4. Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut, hasil nilai *cluster* 1 (satu) iterasi pertama dapat dilihat pada tabel 6.

Tabel 6 Hasil Nilai *Cluster* 1 Iterasi Pertama

Cluster 1			
No	Lokasi	Pemakaian (Jam)	Pemakaian (Ltr)
1	Binjai	1	1
2	Cinta Damai	1	1
3	Delitua	2	1
4	Galang	2	1
5	Kuala	2	1
6	Kuala Putri	2	1
7	Lubuk Pakam	2	1
8	Medan Centrum	1	1
9	Padang Bulan	1	1
10	Pangkalan Brandan	1	1
11	Pangkalan Susu	3	1
12	Perbaungan	2	1
13	Percut	1	1
14	Pulo Brayan	1	1
15	Simpang Limun	1	1
16	Stabat	2	1
17	Sukaramai	1	1
18	Tanjung Pura	3	1
19	Tanjung Morawa	1	1
20	Tanjung Mulia	2	1
21	Tembung	1	1
22	Tuntungan	1	1

$$\text{Centroid} - 1 = \left(\frac{34}{22}, \frac{22}{22} \right) \\ = (1.5 \mid 1)$$

Hasil nilai *cluster* 2 (dua) iterasi pertama dapat dilihat pada tabel 7.

Tabel 7 Hasil Nilai *Cluster* 2 Iterasi Pertama

Cluster 1			
No	Lokasi	Pemakaian (Jam)	Pemakaian (Ltr)
1	Tembung Japati	3	6

$$\text{Centroid} - 2 = \left(\frac{3}{1}, \frac{6}{1} \right) \\ = (3 \mid 6)$$

Untuk *cluster* 2, nilai tiak dibagi karena hanya memiliki 1 nilai, Adapun titik pusat baru dapat dilihat seperti tabel 8.

Tabel 8 Hasil Nilai Titik Pusat Baru setiap *Cluster*

Titik Pusat Baru	Pemakaian (Jam)	Pemakaian (Ltr)
C1	1.5	1
C2	3	6

4. Hitung ulang

Lakukan kembali perhitungan sebelumnya dengan menggunakan *centroid* baru yaitu pada tabel 9. Perhitungan akan menghasilkan beberapa iterasi hingga 2(dua) iterasi terakhir memiliki hasil *cluster* yang sama, maka perhitungan dapat dihentikan.

Tabel 9 Hasil Analisa Cluster Iterasi 2

No	Lokasi	Jarak ke		Jarak Terdekat
		C1	C2	
1	Binjai	0.95	5.38	C1
2	Cinta Damai	0.59	5.38	C1
3	Delitua	0.05	5.09	C1
4	Galang	0.05	5.09	C1
5	Kuala	0.05	5.09	C1
6	Kuala Putri	0.05	5.09	C1
7	Lubuk Pakam	0.05	5.09	C1
8	Medan Centrum	0.95	5.38	C1
9	Padang Bulan	0.95	5.38	C1
10	Pangkalan Brandan	0.95	5.38	C1
11	Pangkalan Susu	1.05	5	C1
12	Perbaungan	0.05	5.09	C1
13	Percut	0.5	5.38	C1
14	Pulo Brayon	0.95	5.38	C1
15	Simpang Limun	0.95	5.38	C1
16	Stabat	0.95	5.09	C1
17	Sukaramai	0.95	5.38	C1
18	Tanjung Pura	1.05	5	C1
19	Tanjung Morawa	0.95	5.38	C1
20	Tanjung Mulia	0.05	5.09	C1
21	Tembung	0.5	5.38	C1
22	Tembung Japati	5.10	0	C2
23	Tuntungan	0.95	5.38	C1

Dalam penelitian ini, iterasi clustering terjadi sebanyak 2 kali iterasi. Pada iterasi ke - 2 ini, titik pusat dari setiap cluster sudah tidak berubah dan tidak ada lagi data yang berpindah dari satu cluster ke cluster yang lain.

3.4 Penerapan K-Means dengan RapidMiner

Data genset perusahaan, selanjutnya di *import* ke *rapidminer*. Pada saat proses *import* data hal yang perlu diperhatikan adalah tipe data. Tipe data yang dapat diolah menggunakan algoritma *k - means* adalah integer. Gambar 2 merupakan hasil konversi data aset pada aplikasi rapid miner.

Row No.	Lokasi	Pemakaian (...)	Pemakaian (...)
1	Binjai	1.580	66.337
2	Cinta Damai	1.400	74.643
3	Delitua	15.740	64.930
4	Galang	19.240	112.254
5	Kuala	17.200	108.067
6	Kuala Putri	17.620	240.591
7	Lubuk Pakam	17.460	350.236
8	Medan Centr...	2.280	593.098
9	Padang Bulan	7.780	212.615
10	Pangkalan Br...	10.020	97.024
11	Pangkalan S...	26.690	166.732
12	Perbaungan	19.180	119.520
13	Percut	10.640	46.923
14	Pulo Brayon	5.330	391.847
15	Simpang Lim...	3.200	109.778
16	Stabat	17.910	306.471

Gambar 2 Konversi Data Perusahaan Menggunakan *RapidMiner*

Data yang telah dikonversi pada *rapidminer*, kemudian dilakukan proses penerapan algoritma *k - means*. Gambar 4.2 adalah proses penerapan algoritma *k - means* pada *rapidminer*. Langkah yang dilakukan dengan cara mengkoneksikan modul-modul, seperti:

1. Retrieve data excel

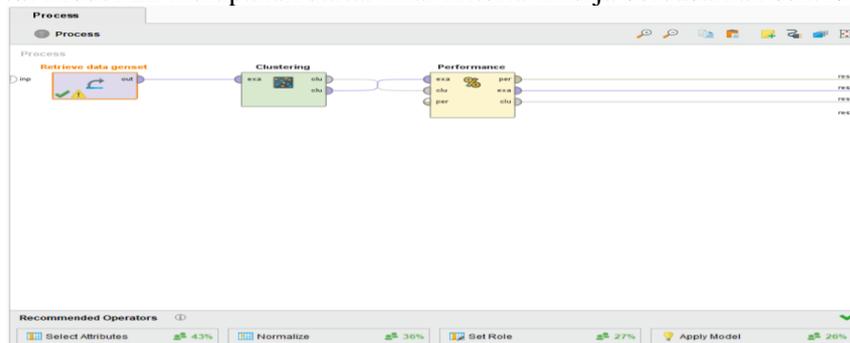
Merupakan modul membaca hasil import data dari *file excel* yang telah disiapkan sebelumnya.

2. Clustering

Merupakan modul yang berfungsi untuk pengelompokan dengan menggunakan algoritma data mining *k-means*.

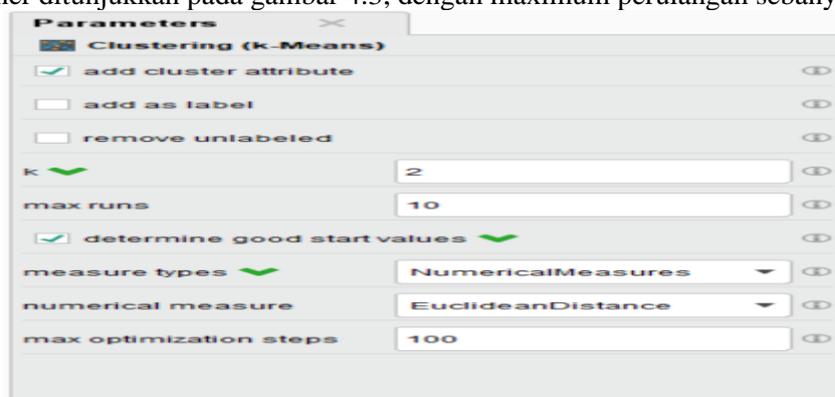
3. Performance

Adalah modul yang berfungsi untuk evaluasi kinerja algoritma pengelompokan berbasis *centroid*. Modul ini merupakan daftar nilai kriteria kinerja berdasarkan *centroid* klaster.



Gambar 3 Proses Penerapan Algoritma K-Means Pada Data Genset pada RapidMiner

Seperti yang dijelaskan sebelumnya, bahwa pengetahuan yang ingin didapatkan adalah mengetahui tentang informasi data genset pada setiap STO berdasarkan penggunaan jam operasi genset tertinggi dan terendah. Dari data tersebut tersebut, maka kriteria yang ingin didapatkan adalah 2. Untuk itu setting pada rapidminer clustering sama dengan 2 (dua) $\rightarrow K = 2$. Setting jumlah cluster pada rapidminer ditunjukkan pada gambar 4.3, dengan maximum perulangan sebanyak 10 kali



Gambar 4 Setting Jumlah Centroid pada RapidMiner

Dari hasil proses penerapan algoritma k-means pada data genset dengan menggunakan rapidminer, diperoleh nilai *centroid* yang ditunjukkan pada gambar 5.

Row No.	id	cluster	Lokasi	Pemakaian (...)	Pemakaian (...)
1	1	cluster_0	Binjai	1.580	66.337
2	2	cluster_0	Cinta Damai	1.400	74.643
3	3	cluster_0	Delitua	15.740	64.930
4	4	cluster_0	Galang	19.240	112.254
5	5	cluster_0	Kuala	17.200	108.067
6	6	cluster_0	Kuala Putri	17.620	240.591
7	7	cluster_0	Lubuk Pakam	17.460	350.236
8	8	cluster_0	Medan Centr...	2.280	593.098
9	9	cluster_0	Padang Bulan	7.780	212.615
10	10	cluster_0	Pangkalan Br...	10.020	97.024
11	11	cluster_0	Pangkalan S...	26.690	166.732
12	12	cluster_0	Perbaungan	19.180	119.520
13	13	cluster_0	Percut	10.640	46.923
14	14	cluster_0	Pulo Brayan	5.330	391.847
15	15	cluster_0	Simpang Lim...	3.200	109.778

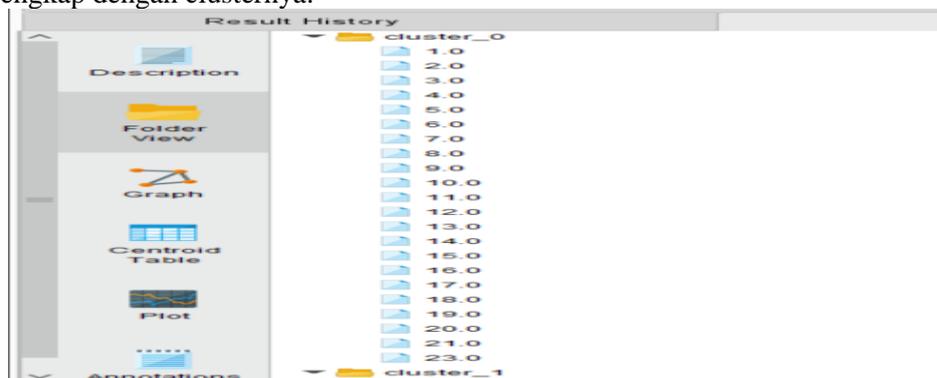
Gambar 5 Visualisasi Data Awal Dengan Data View

Dari 2 (dua) *cluster* yang ditentukan, data genset kemudian dikelompokkan sesuai hasil perhitungan *k-means*. Gambar 6 menunjukkan bahwa cluster 0 (penggunaan jam operasi genset terendah) terdapat di 22 STO, cluster 1 (penggunaan jam operasi genset tertinggi) terdapat 1 STO .



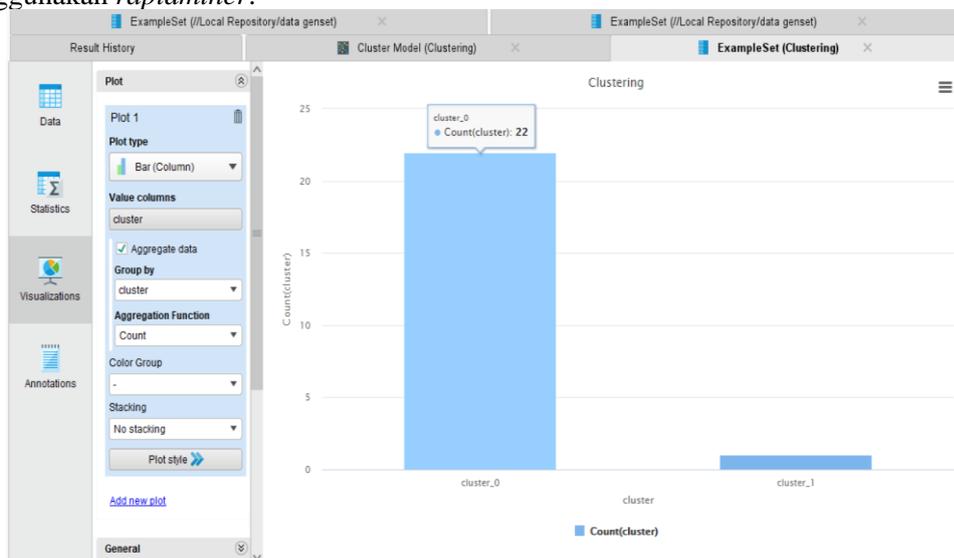
Gambar 6 Hasil Proses *Cluster Model* pada *RapidMiner*

Data yang masuk dalam penggunaan jam operasi genset tertinggi diantaranya adalah STO dan Tembung Japati .Gambar 7 bagian dari untuk menampilkan basis data yang telah diolah secara keseluruhan lengkap dengan clusternya.



Gambar 7 Hasil Proses dalam Folder View Pada *RapidMiner*

Gambar 8 menunjukkan visualisasi *clustering* data genset menggunakan algoritma *k-means* dengan menggunakan *rapidminer*.



Gambar 8 Visualisasi *Cluster K-Means* menggunakan *RapidMiner*

Nilai Davies Bouldin yang dihasilkan dari proses validitas kluster sebesar 0.036, dari dua kluster yang dihasilkan. Semakin kecil nilai validitas kluster yang dihasilkan maka kluster tersebut semakin baik. Gambar 9 merupakan nilai Davies Bouldin hasil proses kluster pada Rapidminer.



Gambar 9 Nilai Davies Bouldin menggunakan Rapidminer

3 KESIMPULAN

Berdasarkan dari hasil penelitian data aset genset, maka penulis menarik beberapa kesimpulan diantaranya sebagai berikut;

1. Pengujian yang dilakukan dalam penelitian ini, iterasi clustering pada data aset genset terjadi sebanyak 2 kali iterasi.
2. Data tersebut diolah menggunakan rapidminer untuk ditentukan nilai dalam centroid dalam 2 cluster yaitu cluster penggunaan jam operasi genset tertinggi(C1) dan cluster penggunaan jam operasi rendah(C2).
3. Dari 23 data lokasi genset berdasarkan pemakaian jam operasi dari genset dan pemakaian solar perliternya dapat diketahui, penggunaan jam operasi genset tertinggi yakni STO Tembung Japati, 22 STO cluster penggunaan jam operasi genset terendah.
4. Nilai Davies Bouldin yang dihasilkan dari proses validitas cluster sebesar 0.036.

REFERENSI

- [1] H. G. Simanullang, A. P. Silalahi, D. Sartika, and U. M. Indonesia, "PREDIKSI JUMLAH PASIEN COVID-19 DI INDONESIA MENGGUNAKAN LEAST SQUARE METHOD BERBASIS ANDROID," vol. 14, no. 1, pp. 86–93, 2022.
- [2] Hajar, S., Novany, A. A., Windarto, A. P., Wanto, A., & Irawan, E. (2020, February). Penerapan K-
- [3] Means Clustering pada ekspor minyak kelapa sawit menurut negara tujuan. In *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)* (Vol. 1, No. 1, pp. 314-318).
- [4] Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan. *Jurnal Tekno Kompak*, 15(2), 25-36.
- [5] Sianipar, K. D. R., Siahaan, S. W., Siregar, M., Zer, F. I. R., & Hartama, D. (2020). Penerapan algoritma k-means dalam menentukan tingkat kepuasan pembelajaran online pada masa pandemi covid-19. *JurTI (Jurnal Teknologi Informasi)*, 4(1), 101-105.
- [6] Wahyuni, W. A., & Saepudin, S. (2021, September). PENERAPAN DATA MINING CLUSTERING UNTUK MENGELOMPOKKAN BERBAGAI JENIS MERK MESIN CUCI. In *Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra* (Vol. 1, No. 01, pp. 306-313).
- [7] Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan. *Jurnal Tekno Kompak*, 15(2), 25-36.