

# ***Re Arrangement Range Pemakaian Gas Pelanggan PGN pada Sub-produk Bronze 1 dan Bronze 2***

**Aditya Pratama Juliyawan**

Matematika, Universitas Negeri Semarang  
Sekaran, Kec. Gn. Pati, Kota Semarang, Jawa Tengah 50229, Indonesia  
*adityajuliyawan@gmail.com*

## **ABSTRAK**

Pemakaian gas pada pelanggan industri memainkan peran penting dalam menentukan efisiensi dan akurasi kebijakan alokasi energi. Namun, batas *range* yang bersifat tetap sering kali tidak mencerminkan pola konsumsi aktual. Penelitian ini bertujuan untuk *re arrangement range* volume pelanggan sub-produk Bronze 1 dan Bronze 2 dengan pendekatan berbasis data. Metode klusterisasi *K-Means*, *Gaussian Mixture Model* (GMM), dan klasifikasi *Jenks Natural Breaks* digunakan untuk menghasilkan batas volume yang lebih relevan, dengan jumlah kluster ditentukan melalui Elbow dan *Silhouette Score*. Validasi dilakukan dengan menganalisis distribusi tingkat penggunaan terhadap volume kontrak, menunjukkan bahwa pelanggan dengan volume lebih kecil seperti yang terjadi pada *range* 1.001-4500 m<sup>3</sup>/bulan, dan 351-1000 MMBtu/bulan, memiliki efisiensi penggunaan lebih tinggi. Hasil klusterisasi juga mengarah pada pembentukan sub-produk baru yang lebih mencerminkan perilaku konsumsi, seperti Bronze 1a dan Bronze 1b serta Bronze 2a dan Bronze 2b. Temuan ini memberikan dasar yang lebih akurat untuk pengambilan keputusan berbasis data dalam pengelolaan kontrak dan strategi pemasaran energi.

**Kata kunci :** *Klaster, Klasifikasi, K-Means, Gaussian Mixture Model, Jenks Natural Breaks Optimization*

## **1. PENDAHULUAN**

PT Perusahaan Gas Negara Tbk (PGN) merupakan perusahaan nasional terbesar di bidang transmisi dan distribusi gas bumi di Indonesia, dengan peran penting dalam pemenuhan kebutuhan gas domestik [1]. Dalam industri energi, khususnya sektor gas, klasifikasi produk pelanggan menjadi pendekatan utama dalam analisis bisnis dan pemasaran. Melalui metode *unsupervised learning* berbasis statistik dan matematika, perusahaan dapat menemukan pola konsumsi dan mengelompokkan pelanggan ke dalam *range* homogen yang merepresentasikan karakteristik serupa [2].

PGN telah menetapkan produk dan sub-produk pelanggan industri berdasarkan volume pemakaian gas bumi, baik harian, bulanan, maupun tahunan. Salah satunya adalah produk bulanan Bronze. Namun, pada sub-produk Bronze 1, dengan *range* pemakaian antara 1.001 - 10.000 m<sup>3</sup>/bulan, serta sub-produk Bronze 2 dengan *range* nilai antara 350 – 1750 MMBtu/bulan, dinilai terlalu luas dan kurang mencerminkan kebutuhan serta karakteristik pelanggan secara proporsional. *Range* ini dapat menyebabkan ketidak efisienan dalam pelayanan, pengelolaan produk langganan, serta penetapan strategi layanan yang tepat.

Permasalahan terkait *range* klasifikasi pelanggan yang terlalu lebar bukanlah hal baru di industri. Seiring perkembangan teknologi analisis data, pendekatan berbasis data historis kini banyak digunakan untuk menyempurnakan *range* dan meningkatkan akurasi klasifikasi [3]. Beberapa penelitian menekankan pentingnya pemanfaatan data historis volume pemakaian sebagai dasar penentuan *range* yang lebih akurat, karena pendekatan ini tidak hanya meningkatkan efisiensi distribusi, tetapi juga

mendukung strategi harga yang lebih adil dan tepat sasaran [4].

Berdasarkan permasalahan tersebut, artikel ini bertujuan untuk merumuskan *range* baru dalam sub-produk Bronze 1 dan Bronze 2 yang dibagi menjadi beberapa sub-produk dengan pendekatan analisis statistik eksploratif dan algoritma seperti *K-Means Clustering*, *Gaussian Mixture Model* (GMM), dan *Jenks Natural Breaks Optimization*. Hasil dari *re arrangement range* ini diharapkan dapat menghasilkan struktur klasifikasi yang lebih optimal dari sisi statistik maupun operasional, serta dapat dijadikan acuan dalam pengambilan keputusan strategis di internal PGN, khususnya dalam hal penetapan tarif, pengembangan layanan pelanggan, dan optimalisasi distribusi gas bumi.

Metode *re arrangement range* pelanggan berbasis data yang umum digunakan dalam analisis pelanggan mencakup algoritma *K-Means*. *K-Means* merupakan metode *non-hierarchical clustering* yang membagi data ke dalam sejumlah kluster berdasarkan tingkat kemiripan, dengan mengoptimalkan jarak antar pusat kluster dan titik data [5]. Sementara itu, GMM merupakan metode klusterisasi berbasis probabilistik yang mengasumsikan bahwa data berasal dari campuran beberapa distribusi normal. Metode ini memberikan fleksibilitas dalam menangani data yang memiliki distribusi yang tumpang tindih dan tidak linear [6]. Di sisi lain, *Jenks Natural Breaks Optimization* adalah metode klasifikasi yang digunakan untuk meminimalkan variasi dalam setiap kelompok dan memaksimalkan perbedaan antar kelompok. Metode ini sangat efektif untuk *re arrangement range* data dengan sebaran yang tidak

Dalam studi sebelumnya, pengelompokan pelanggan berbasis *unsupervised learning* telah terbukti efektif dalam menganalisis pola penggunaan dan *re arrangement range*. Salah satu penelitian yang menggunakan pendekatan ini adalah oleh Wang J. (2023), yang memanfaatkan algoritma K-Means dan Gaussian Mixture Model (GMM) untuk mengklasifikasikan pengguna berdasarkan data historis. [8].

### 3.1 Pengumpulan Data

Tahap pengumpulan data dilakukan dengan mengambil data historis pemakaian gas bumi dari pelanggan industri yang tercatat oleh PGN dan diambil dari periode 2022 hingga 2024. Data ini terdiri dari sejumlah variabel penting yang mencakup SBU, Kode Area, No Referensi, Nama Akun, Sektor Industri, Periode, Volume Terukur (dalam m<sup>3</sup>/bulan dan MMBtu/bulan), serta Batas Minimum dan Batas Maximum. Data ini berfungsi sebagai dasar untuk proses eksplorasi, *re arrangement range*, dan penentuan sub-produk pelanggan.

### 3.2 Pengolahan Data

Pengolahan data dilakukan dalam enam tahap :

#### 1) Preprocessing Data

Tahap *preprocessing* bertujuan untuk memastikan data yang digunakan dalam analisis berada dalam kondisi bersih, konsisten, dan siap olah. Proses ini mencakup pengecekan data duplikat, penanganan pada data yang hilang (*missing value*), standarisasi format data, reduksi data dan pembersihan data agar data menjadi siap untuk diolah. Selain itu, dilakukan pemilihan variabel dan penambahan variabel baru yang relevan untuk analisis dan normalisasi numerik agar metode klasterisasi dapat bekerja optimal.

#### 2) Eksplorasi Pola dan Distribusi Data

Eksplorasi data dilakukan untuk mengidentifikasi distribusi, pola, dan karakteristik dari volume pemakaian gas oleh pelanggan industri. Analisis ini meliputi visualisasi distribusi volume per periode waktu, identifikasi tren konsumsi berdasarkan sektor industri dan wilayah, serta analisis statistik deskriptif untuk mendeskripsikan kecenderungan umum dalam data. Informasi ini menjadi landasan awal dalam menentukan pendekatan klasterisasi yang paling sesuai.

#### 3) Pencarian Jumlah Klaster

Penentuan jumlah klaster optimal merupakan langkah penting dalam klasterisasi. Metode Elbow digunakan untuk mengidentifikasi titik dimana penambahan jumlah klasterisasi tidak lagi menghasilkan penurunan signifikan (titik “elbow”) terhadap nilai WCSS (*Within-Cluster sum of squares*) [13]. Nilai WCSS dihitung menggunakan rumus :

$$WCSS = \sum_{i=1}^K \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (6)$$

Setelah memperoleh jumlah klaster yang optimal, diperoleh informasi hasil klastering berdasarkan klaster yang optimal [14]. Disisi lain, *Silhouette Score* mengukur kualitas klaster dengan membandingkan kedekatan suatu data terhadap klaster tempat ia berada dan klaster terdekat lainnya, menggunakan rumus:

$$s(x_i) = \frac{b(x_i) - a(x_i)}{\max\{b(x_i), a(x_i)\}} \quad (7)$$

Dimana  $x_i$  adalah sebuah elemen di klaster  $\pi_k$ , dan  $a(x_i)$  adalah jarak rata rata dari  $x_i$  untuk semua elemen lain di klaster  $\pi_k$ .

$$b = \min\{d_l(x_i)\}, \text{ dimana } l \neq k. \quad (8)$$

Dimana  $d_l(x_i)$  adalah rata rata jarak dari  $x_i$  ke semua titik di klaster  $\pi_l$  untuk  $l \neq k$  [11]. Hasil dari tahap ini akan digunakan sebagai acuan dalam penerapan algoritma klasterisasi seperti *K-Means*, GMM dan JNB pada tahap selanjutnya.

#### 4) Melakukan Re Arrangement Range

*Re Arrangement range* dilakukan terhadap data volume konsumsi gas pelanggan menggunakan tiga metode utama, yaitu *K-Means Clustering*, *Gaussian Mixture Model* (GMM), dan *Jenks Natural Breaks Optimization*. *K-Means* mengelompokkan data berdasarkan kedekatan terhadap pusat klaster, dengan tujuan meminimalkan jarak kuadrat antar data terhadap *centroid* klaster :

$$J = \sum_{k=1}^K \sum_{x_i \in \pi_k} \|x_i - \mu_k\|^2 \quad (9)$$

Dimana  $\pi_k$  adalah klaster k,  $\mu_k$  adalah pusat klaster dari klaster k, dan  $\| \cdot \|$  adalah jarak Euclidian. GMM digunakan sebagai pendekatan alternatif untuk menangkap distribusi data yang tidak linier dan tidak homogen. Tidak seperti *K-Means* [11], GMM mengasumsikan bahwa data berasal dari kombinasi beberapa distribusi normal dan menggunakan pendekatan probabilistik untuk menentukan keanggotaan klaster tiap data poin [8]. Sementara itu, *Jenks Natural Breaks* merupakan metode berbasis optimasi variansi, yang bertujuan untuk meminimalkan variansi dalam kelompok dan memaksimalkan perbedaan antar kelompok. Metode ini sangat efektif ketika data bersifat satu dimensi dan memiliki distribusi tidak merata, seperti volume konsumsi gas industri [12].

#### 5) Re Arrangement Sub-Produk Baru

Hasil sub-produk baru divisualisasikan menggunakan berbagai teknik untuk mempermudah interpretasi dan analisis tiap klaster. *Scatter plot* digunakan untuk menunjukkan distribusi dan pemisahan antar klaster berdasarkan volume konsumsi. *Bar chart* dan *pie chart* digunakan untuk menampilkan proporsi pelanggan dalam setiap *range*. Sementara itu, tabel statistik menyajikan informasi deskriptif seperti volume pemakaian per klaster, batas minimal dan maksimal, serta persentase pelanggan yang ada di dalam *range* tersebut.

### 3.3 Metode Analisis

Berikut merupakan analisis pemilihan metode beserta *tools* yang digunakan dalam penelitian ini.

#### 1) Teknik Analisis Klaster

Penelitian ini menggunakan pendekatan *unsupervised learning* untuk melakukan *re*

*arrangement* pelanggan produk Bronze, khususnya sub-produk Bronze 1 dan Bronze 2, berdasarkan data volume pemakaian aktual. Tiga algoritma digunakan dalam analisis ini, yaitu *K-Means Clustering*, *Gaussian Mixture Model (GMM)*, dan *Jenks Natural Breaks Optimization*. Ketiga metode ini dipilih karena masing-masing merepresentasikan pendekatan yang berbeda terhadap struktur data. *K-Means* merupakan algoritma berbasis partisi yang bekerja dengan mengelompokkan data berdasarkan kedekatan terhadap pusat kluster, sehingga efektif digunakan pada data yang memiliki distribusi seragam. Sementara itu, GMM menggunakan pendekatan probabilistik yang mampu menangkap kompleksitas distribusi data yang saling tumpang tindih, sehingga lebih fleksibel dalam menangani keragaman pola konsumsi. *Jenks Natural Breaks*, di sisi lain, memfokuskan pada identifikasi batas alami dalam distribusi data dengan memaksimalkan perbedaan antar kelompok dan meminimalkan varians dalam satu kelompok, sehingga cocok diterapkan pada data yang memiliki distribusi tidak merata. Penggunaan ketiga algoritma ini bertujuan bukan untuk menentukan metode terbaik, melainkan untuk membandingkan hasil *range* yang diperoleh guna mengevaluasi kesesuaian masing-masing pendekatan terhadap pola konsumsi aktual pelanggan.

- 2) Penentuan Jumlah Kluster Optimal  
Penentuan jumlah kluster optimal dilakukan dengan memanfaatkan metode Elbow dan *Silhouette Score*. Metode Elbow berguna untuk menentukan titik optimal dimana penambahan jumlah kluster tidak lagi memberikan penurunan signifikan pada nilai *within-Cluster Sum of Squares (WCSS)*, sementara itu *Silhouette Score* digunakan untuk mengukur seberapa baik objek objek dikelompokkan dalam kluster yang sesuai.
- 3) *Tools dan Software Analisis*  
Seluruh proses pengolahan dan analisis data dilakukan menggunakan bahasa pemrograman Python dengan *library* seperti pandas, matplotlib, seaborn dan scikit-learn. Python dipilih karena memiliki *library* standar yang sangat besar dan kuat yang dapat digunakan untuk menganalisis dan memvisualisasikan data [15].

#### 4. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dari analisis kluster yang telah dilakukan untuk *re arrangement range* pemakaian pelanggan sub-produk Bronze 1 dan sub-produk Bronze 2. Setiap tahapan mulai dari deskripsi data, *preprocessing* data, hingga hasil klustering dan klasifikasi dengan algoritma *K-Means*, *Gaussian Mixture dan Jenks Natural Breaks*, akan diuraikan secara rinci.

##### 4.1 Deskripsi Data

Data dalam penelitian ini adalah data realisasi pelanggan sub-produk Bronze 1 dan Bronze 2, yang mencakup informasi volume pemakaian gas pelanggan (dalam satuan m<sup>3</sup>/bulan dan MMBtu/bulan), serta batas minimum dan maksimum selama periode tertentu, dengan tabahan minimal kontrak dan maksimal kontrak setiap periode. Data terdiri dari 118.228 baris dan 19 kolom, dengan variable utama berupa, ID pelanggan, SBU, periode waktu, jumlah pemakaian aktual, dan batas minimal serta maksimal kontrak, data ini digunakan sebagai dasar untuk analisis distribusi dan *re arrangement range* pemakaian agar lebih mencerminkan kondisi di lapangan.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 118228 entries, 0 to 118227
Data columns (total 19 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   SBU                  118228 non-null object
 1   Kode Area            118227 non-null object
 2   No Ref               118227 non-null object
 3   Account Name         118227 non-null object
 4   Sektor Industri     118227 non-null object
 5   Periode              118227 non-null object
 6   Sum of Terukur m3    118228 non-null object
 7   Sum of Terukur mmbtu 118227 non-null object
 8   Sum of Batas Min     118228 non-null object
 9   Sum of Batas Max     118228 non-null object
10   Kode Area            118227 non-null object
11   No Ref               118227 non-null object
12   Account Name         118227 non-null object
13   Sektor Industri     118227 non-null object
14   Periode              118227 non-null object
15   Terukur m3          118227 non-null object
16   Terukur mmbtu        118226 non-null object
17   Sum of Batas Min     118227 non-null object
18   Sum of Batas Max     118227 non-null object
dtypes: object(19)
memory usage: 17.1+ MB
```

Gambar 3. Informasi Data

##### 4.2 Preprocessing Data

Pada tahap ini dilakukan beberapa langkah utama dalam *preprocessing* data diantaranya adalah sebagai berikut.

- 1) Pembersihan Data  
Langkah pertama dalam *preprocessing* adalah menghapus kolom yang terduplikasi untuk memastikan tidak ada informasi yang berulang dan memengaruhi hasil analisis. Setelah itu, data diperiksa untuk menemukan *missing value*. Lalu dihilangkan agar dataset yang dianalisis bersih dan valid.
- 2) Pengolahan dan Penambahan Kolom  
Selanjutnya dilakukan penambahan kolom baru berupa kategori pelanggan dengan isi kolom terdiri dari Bronze 1 dan Bronze 2. Kategori ini ditentukan berdasarkan nilai pada kolom *Sum of Batas Min* dan *Sum of Batas Max*, dimana jika nilai batas minimum adalah  $1000 \leq x \leq 1601$  dan batas maksimum adalah  $6000 \leq y \leq 16500$  maka akan masuk ke sub-produk Bronze 1. Sementara itu, jika nilai batas minimum  $x \leq 350$  dan batas maksimum  $y \leq 4000$  data akan masuk ke sub-produk Bronze 2.

```

<class 'pandas.core.frame.DataFrame'>
Index: 111727 entries, 0 to 118226
Data columns (total 11 columns):
#   Column              Non-Null Count  Dtype  
---  -
0   SBU                  111727 non-null  object 
1   Kode Area            111727 non-null  int64  
2   No Ref               111727 non-null  int64  
3   Account Name         111727 non-null  object 
4   Sektor Industri      111727 non-null  object 
5   Periode              111727 non-null  datetime64[ns]
6   Terukur m3          111727 non-null  float64 
7   Terukur mmbtu        111726 non-null  float64 
8   Batas Min            111727 non-null  float64 
9   Batas Max            111727 non-null  float64 
10  segmentasi           111727 non-null  object 
dtypes: datetime64[ns](1), float64(4), int64(2), object(4)
memory usage: 10.2+ MB

```

Gambar 4. Hasil Pengolahan dan Penambahan Kolom

### 3) Pemisahan Dataset

Setelah data bersih dan siap digunakan, dataset kemudian dibagi menjadi dua sub-produk, yaitu Bronze 1 dan Bronze 2, proses ini dilakukan agar proses klasterisasi dapat dilakukan secara lebih terfokus sesuai dengan karakteristik masing-masing. Selain itu, data juga dipisahkan berdasarkan periode waktu yaitu antara tahun 2022 samapi dengan 2024, untuk melihat dinamika atau perbandingan yang terjadi antar periode.

```

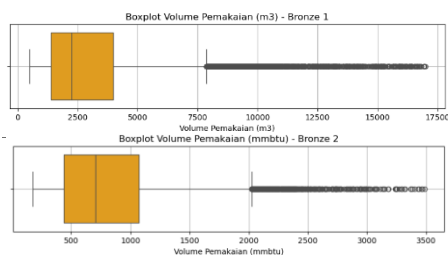
df_bronze1.to_excel('data_cleanedbronze1_PGN.xlsx', index=False)
df_bronze2.to_excel('data_cleanedbronze2_PGN.xlsx', index=False)

```

Gambar 5. Pembagian Data Menjadi 2 Sub-produk

### 4) Penanganan Outlier

Untuk menjaga konsistensi dan akurasi model, baris-baris data yang mengandung pencilan (*outlier*) diidentifikasi dan dihapus. Identifikasi pencilan dilakukan dengan metode visualisasi seperti boxplot untuk membantu melihat data yang menyimpang secara ekstrem dari distribusi normal. Penanganan outlier dilakukan dengan kriteria bahwa data akan dibersihkan jika terdapat volume pemakaian gas yang melebihi dua kali lipat dari nilai Batas Max atau lebih kecil dari setengah Batas Min.



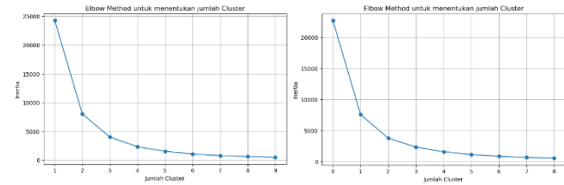
Gambar 6. Box Plot Bronze 1 dan Bronze 2

### 4.3 Penentuan Jumlah Klaster Optimal

Setelah data di bersihkan dan penanganan *outlier*, selanjutnya adalah menentukan jumlah klaster yang optimal menggunakan metode Elbow dan *Silhouette Score* dengan menggunakan *library scikit-learn*.

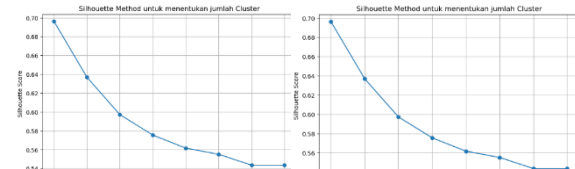
Dapat dilihat pada Gambar 7 dan 8, bahwa untuk Metode Elbow hasil menunjukkan jumlah klater optimal adalah k=3 dan meskipun k=2 juga masih memungkinkan karena perbedaan nilai tidak signifikan untuk sub-produk *Bronze 1*, Sementara itu,

untuk sub-produk *Bronze 2*, hasil menunjukkan bahwa K = 2 merupakan pilihan yang paling optimal.



Gambar 7. Metodew Elbow untuk Bronze 1 dan Bronze 2

Sebagai pembanding, nilai *Silhouette Score* memberikan ukuran yang lebih objektif karena bersifat kuantitatif dan menggambarkan seberapa baik setiap data berada dalam klaster yang sesuai. Hasil menunjukkan bahwa nilai *Silhouette Score* untuk Bronze 1 sebesar 0.69 dan untuk Bronze 2 sebesar 0.62. Nilai ini menunjukkan bahwa klaster yang terbentuk sudah cukup baik, dengan batas antar kelompok yang relatif jelas. Namun demikian, skor tersebut juga mengindikasikan adanya sebagian kecil data yang berada di area perbatasan antar klaster.



Gambar 8. Metode *Silhouette* untuk Bronze 1 dan Bronze 2

### 4.4 Klastering dengan K-Means dan GMM

Setelah didapatkan jumlah klaster yang optimal pada metode *Silhouette*, maka selanjutnya adalah melakukan proses klasterisasi. Penelitian ini menggunakan dua metode utama, yaitu *K-Means* dan *Gaussian Mixture Model* (GMM).

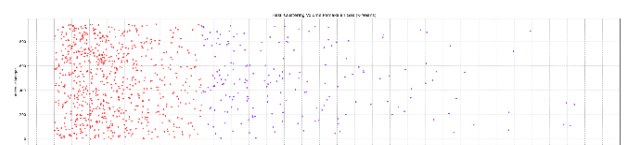
#### 1) Algoritma *K-Means*

##### a. Bronze 1

Tabel 1. Statistik Deskriptif Volume *K-Means* Bronze 1

K-Means	Count	Mean	STD	MIN	50%	Max
0	752	2110,24	1053,45	505,46	1894,71	4692,50
1	191	7316,42	2395,67	4715,68	6617,46	15208

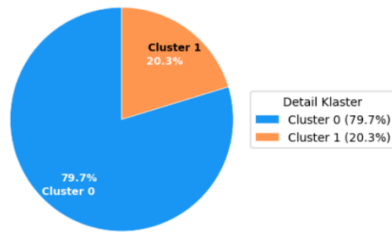
Pemodelan *K-Means* menghasilkan dua klaster pelanggan berdasarkan pola pemakaian gas. Klaster 0 mencakup 752 pelanggan dengan rata rata pemakaian gas sebesar 2.110,24 m<sup>3</sup>/bulan, range 505,46 – 4.692,50 m<sup>3</sup>/bulan, dan standar deviasi sebesar 1.053,45. Klaster 1 mencakup 191 pelanggan dengan rata rata pemakaian 7.316,42 m<sup>3</sup>/bulan, range 4.715,68–15.208 m<sup>3</sup>/bulan, dan standar deviasi 2.395,67. Perbedaan ini menunjukkan bahwa model berhasil memisahkan pelanggan menjadi dua kelompok utama: pemakaian rendah-sedang (Klaster 0) dan pemakaian tinggi (Klaster 1).



Gambar 9. Visualisasi *Scatter Plot K-Means* Bronze 1



Pada diagram lingkaran dibawah terdapat sekitar 79.7% atau 752 perusahaan yang masuk pada kluster 0, dan 20,3% atau 191 perusahaan yang masuk kluster 1, berdasarkan rata-rata pemakaian gas per perusahaan.



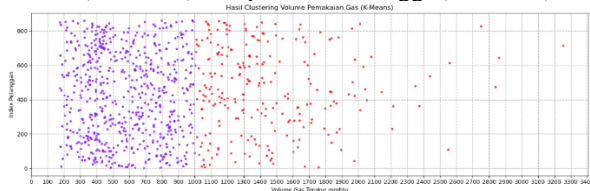
Gambar 10. Persentase Jumlah Perusahaan *K-Means* Bronze 1

#### b. Bronze 2

Tabel 2. Statistik Deskriptif Volume *K-Means* Bronze 2

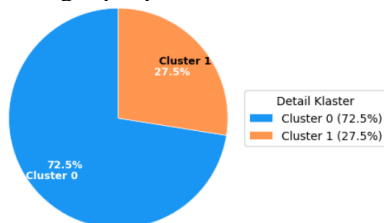
<i>K-Means</i>	Count	Mean	STD	MIN	50%	Max
0	620	576,94	226,18	175,84	563,31	1003,88
1	242	1438,61	1008,81	1149,66	1337,42	3252,80

Kluster 0 mencakup 620 pelanggan dengan rata rata pemakaian gas sebesar 576,94 MMBtu/bulan, *range* 175,84 – 1.003,88 MMBtu/bulan, dan standar deviasi sebesar 226,18. Kluster 1 mencakup 242 pelanggan dengan rata rata pemakaian 1008 MMBtu/bulan, *range* 1.149,66 – 3252,80 MMBtu/bulan, dan standar deviasi 226,18. Perbedaan ini menunjukkan bahwa model berhasil memisahkan pelanggan menjadi dua kelompok utama: pemakaian rendah-sedang (Kluster 0) dan pemakaian tinggi (Kluster 1).



Gambar 11. Visualisasi *Scatter Plot K-Means* Bronze 2

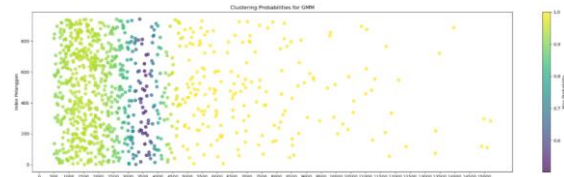
Pada diagram lingkaran dibawah terdapat sekitar 72.5% atau 620 perusahaan yang masuk pada kluster 0, dan 27,5% atau 242 perusahaan yang masuk kluster 1, berdasarkan rata-rata pemakaian gas per perusahaan.



Gambar 12. Persentase Jumlah Perusahaan *K-Means* Bronze 2

## 2) Algoritma *Gaussian Mixture Model*

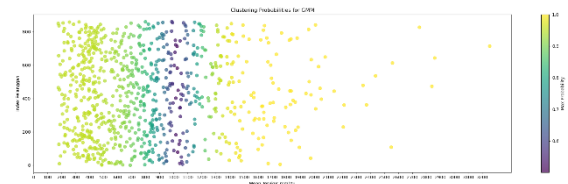
### a. Bronze 1



Gambar 13. Probabilitas klustering untuk GMM Bronze 1

Visualisasi probabilitas kluster GMM menunjukkan adanya zona ambiguitas pada *range* konsumsi rata-rata 3200 – 4150 m<sup>3</sup>/bulan, di mana probabilitas keanggotaan kluster cenderung rendah (sekitar 0.5–0.8). Hal ini mengindikasikan bahwa pelanggan dalam *range* ini tidak secara jelas tergolong ke salah satu kluster, atau berada di wilayah perbatasan antar kluster. Sebaliknya, pelanggan dengan konsumsi di bawah 3000 m<sup>3</sup>/bulan atau di atas 4500 m<sup>3</sup>/bulan menunjukkan probabilitas tinggi (>0.9), menandakan pemisahan kluster yang lebih tegas. Dengan demikian, *range* 3000–4100 m<sup>3</sup>/bulan merupakan area transisi yang perlu perhatian khusus dalam interpretasi *range* dan penentuan kebijakan berbasis kluster.

### b. Bronze 2



Gambar 14. Probabilitas klustering untuk GMM Bronze 2

Visualisasi probabilitas kluster GMM menunjukkan adanya zona ambiguitas pada *range* konsumsi rata-rata 900 – 1100 MMBtu/bulan, di mana probabilitas keanggotaan kluster cenderung rendah (sekitar 0.5–0.8). Hal ini mengindikasikan bahwa pelanggan dalam *range* ini tidak secara jelas tergolong ke salah satu kluster. Dengan demikian, *range* 900–1100 MMBtu/bulan merupakan area transisi yang perlu perhatian khusus dalam interpretasi *range* dan penentuan kebijakan berbasis kluster.

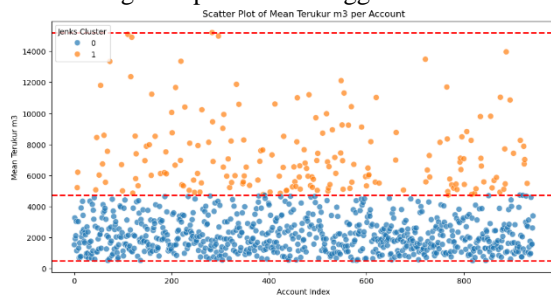
## 4.5 Klasifikasi dengan *Jenks Natural Breaks*

### a. Bronze 1

Tabel 3. Statistik Deskriptif Volume Klasifikasi JNB

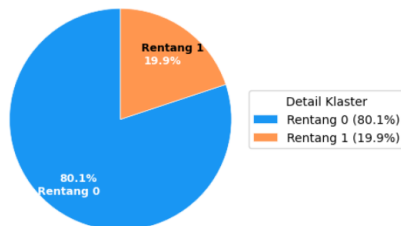
JNB	Count	Mean	STD	MIN	50%	Max
0	755	2120,61	1064,12	505,46	1898,90	4731,06
1	188	7357,83	2391,97	4739,86	6677,82	15208,95

*Range 0* mencakup 755 pelanggan dengan rata-rata pemakaian gas sebesar 2120,61 m<sup>3</sup>/bulan, *range* 505,46 – 4.731,06 m<sup>3</sup>/bulan, dan standar deviasi sebesar 1064,12. *Range 1* mencakup 188 pelanggan dengan rata rata pemakaian 7357,83 m<sup>3</sup>/bulan, *range* 4.739,86 – 15.208,95 m<sup>3</sup>/bulan, dan standar deviasi 2.391,97. Perbedaan ini menunjukkan bahwa model berhasil memisahkan pelanggan menjadi dua kelompok utama: pemakaian rendah–sedang dan pemakaian tinggi.



Gambar 15. Persebaran Klasifikasi untuk JNB Bronze 1

Pada diagram lingkaran dibawah terdapat sekitar 80,1% atau 755 perusahaan yang masuk pada *range 0*, dan 19,9% atau 188 perusahaan yang masuk *range 1*, berdasarkan rata-rata pemakaian gas per perusahaan.



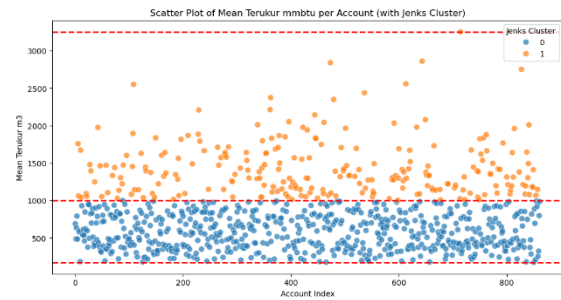
Gambar 16. Persentase Jumlah Perusahaan JNB Bronze 2

#### b. Bronze 2

Tabel 4. Statistik Deskriptif Volume Klasifikasi JNB Bronze 2

JNB	Count	Mean	STD	Min	50%	Max
0	619	576.25	225.71	175.84	561,63	1000,22
1	243	1436.82	376.96	1003.88	1336,13	3252,80

*Range 0* mencakup 619 pelanggan dengan rata-rata pemakaian gas sebesar 576,245 MMBtu/bulan, *range* 175,84 – 1.000,22 MMBtu/bulan, dan standar deviasi sebesar 225,71. *Range 1* mencakup 243 pelanggan dengan rata rata pemakaian 1.436,82 MMBtu/bulan, *range* 1.003,88 – 3.252,80 MMBtu/bulan, dan standar deviasi 376,96. Perbedaan ini menunjukkan bahwa model berhasil memisahkan pelanggan menjadi dua kelompok utama: pemakaian rendah–sedang dan pemakaian tinggi.



Gambar 17. Persebaran Klasifikasi untuk JNB Bronze 2

Pada diagram lingkaran dibawah terdapat sekitar 71,8% atau 619 perusahaan yang masuk pada *range 0*, dan 28,2% atau 243 perusahaan yang masuk *range 1*, berdasarkan rata-rata pemakaian gas per perusahaan.

#### 4.6 Uji Data dan Validasi *Range*

Uji data dan validasi *range* dilakukan untuk memastikan sub-produk pelanggan Bronze 1 dan Bronze 2 menghasilkan volume pemakaian secara akurat. Penentuan jumlah klaster optimal menggunakan metode Elbow dan Silhouette Score menunjukkan bahwa K=2 paling sesuai untuk kedua sub-produk. Selanjutnya, dilakukan klasterisasi menggunakan *K-Means*, *Gaussian Mixture Model* (GMM), dan klasifikasi menggunakan *Jenks Natural Breaks*. Pada sub-produk Bronze 1, hasil ketiga metode menunjukkan variasi, *K-Means* dengan *range* 505 – 4.692 m<sup>3</sup>/bulan, JNB 505 – 4.731 m<sup>3</sup>/bulan, dan GMM mengindikasikan area transisi antara 3.000–4.100 m<sup>3</sup>/bulan. Sementara itu, pada sub-produk Bronze 2 menunjukkan hasil konsisten di kisaran 175 – 1.003 m<sup>3</sup>. Untuk menghasilkan range akhir yang lebih netral dan representatif, digunakan rata-rata batas bawah dan atas dari ketiga metode. Pendekatan ini meminimalkan bias dan mendukung *range* yang lebih akurat berbasis data.

Tabel 5. *Range* Sub-produk Bronze 1 dan Bronze 2

Sub-Produk	Batas Minimal	Batas Maksimal	Persentase Kenaikan
Bronze 1a	1.001 m <sup>3</sup>	4.500 m <sup>3</sup>	349,5%
Bronze 1b	4.501 m <sup>3</sup>	10.000 m <sup>3</sup>	122,2%
Bronze 2a	350 MMBtu	1.000 MMBtu	185,7%
Bronze 2b	1.001 MMBtu	1.750 MMBtu	75%

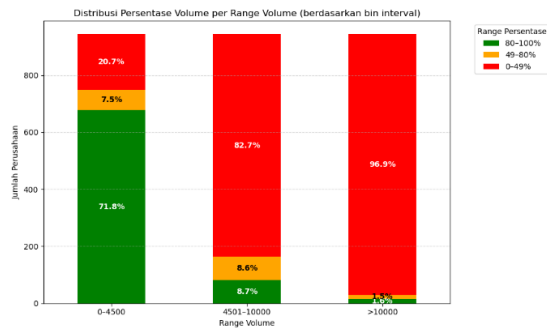
#### a. Bronze 1

Tabel 6. Frekuensi per *Range* Sub-produk Bronze 1

<i>Range</i> (m <sup>3</sup> /bulan)	80-100%	50-80%	0-50%
0-4.500	677 (71,8%)	71 (7,5%)	195 (20,7%)
4.501-10.000	82 (8,7%)	81 (8,6%)	780 (82,7%)
>10.000	15 (1,6%)	14 (1,5%)	914 (96,9%)

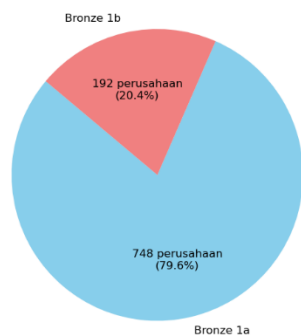
Tabel 6 menyajikan distribusi frekuensi pelanggan Bronze 1 berdasarkan tiga *range* volume yang diuji: 0–4.500 m<sup>3</sup>/bulan, 4.501–10.000 m<sup>3</sup>/bulan, dan gabungan >10.000 m<sup>3</sup>/bulan. Masing-masing *range* dievaluasi berdasarkan persentase pelanggan yang volume

aktualnya berada pada level 80–100%, 50–80%, dan 0–50% dari rentang tersebut.



Gambar 18. Distribusi Volume per Range Bronze 1

Hasilnya menunjukkan bahwa pelanggan dengan volume pemakaian di bawah 4.500 m<sup>3</sup>/bulan cenderung memiliki tingkat utilisasi yang tinggi, dengan mayoritas berada pada kisaran 80–100%. Sebaliknya, pelanggan yang termasuk dalam rentang volume 4.501–10.000 m<sup>3</sup>/bulan sebagian besar hanya menggunakan gas dalam kisaran 0–50% dari total rentang tersebut.



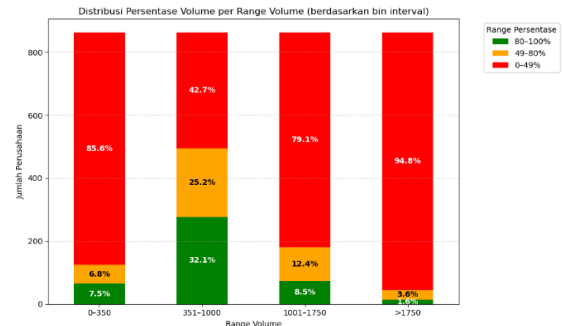
Gambar 19. Distribusi Pelanggan Bronze 1a dan Bronze 1b

## b. Bronze 2

Tabel 7. Frekuensi per Range Sub-produk Bronze 2

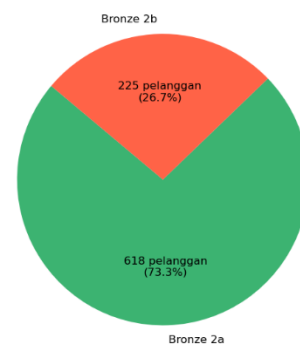
Range (MMBtu/bulan)	80-100%	50-80%	0-50%
0-350	65 (7,5%)	59 (6,8%)	738 (85,6%)
351-1.000	277 (32,1%)	217 (25,2%)	368 (42,7%)
1.001-1.750	73 (8,5%)	107 (12,4%)	682 (79,1%)
>1.750	14 (1,6%)	31 (3,6%)	817 (94,8%)

Tabel 7 menyajikan distribusi frekuensi pelanggan Bronze 2 berdasarkan empat range volume yang diuji: 0–350 MMBtu/bulan, 351–1.000 MMBtu/bulan, 1.001–1.750 MMBtu/bulan dan  $\geq 1.750$  MMBtu/bulan. Masing-masing range dievaluasi berdasarkan persentase pelanggan yang volume aktualnya berada pada level 80–100%, 50–80%, dan 0–50% dari rentang tersebut.



Gambar 20. Distribusi Volume per Range Bronze 2

Hasil analisis menunjukkan bahwa pelanggan dengan volume pemakaian pada rentang 351–1000 MMBtu/bulan cenderung memiliki tingkat utilisasi yang tinggi, dengan mayoritas menggunakan lebih dari 50% dari kontrak volume-nya. Berdasarkan diagram lingkaran, terdapat 618 pelanggan (sekitar 73,3%) yang termasuk dalam sub-produk 2a ( $\leq 1000$  MMBtu/bulan dengan utilisasi  $\geq 50\%$ ), sedangkan 225 pelanggan (sekitar 26,7%) tergolong dalam sub-produk 2b ( $>1000$  MMBtu/bulan dengan utilisasi  $\geq 50\%$ ).



Gambar 21. Distribusi Pelanggan Bronze 2a dan Bronze 2b

## 5. KESIMPULAN DAN SARAN

Penelitian ini berhasil melakukan *re arrangement range* volume gas untuk pelanggan sub-produk Bronze 1 dan Bronze 2 menggunakan pendekatan klusterisasi dan klasifikasi berbasis data. Penentuan jumlah kluster dilakukan melalui metode Elbow dan Silhouette Score, yang menunjukkan hasil optimal  $k=2$  hingga  $k=3$ . Tiga algoritma yaitu K-Means, GMM, dan Jenks Natural Breaks digunakan untuk menghasilkan batas volume, dengan hasil yang kemudian dirata-ratakan untuk memperoleh range akhir yang lebih netral dan representatif. Validasi dilakukan dengan menganalisis distribusi tingkat utilisasi pelanggan terhadap volume kontrak, yang menunjukkan bahwa pelanggan dengan volume lebih kecil seperti sub-produk Bronze 1a dan Bronze 2a, cenderung memiliki efisiensi pemakaian lebih tinggi. Visualisasi melalui bar chart dan pie chart mendukung kategori pelanggan menjadi dua sub-produk berbasis kombinasi volume dan utilisasi (misalnya Bronze 1a dan 1b, serta Bronze 2a dan 2b), sehingga pendekatan ini terbukti efektif dalam memberikan dasar yang lebih akurat dan berbobot



untuk mendukung pengambilan keputusan bisnis terkait pengelolaan volume gas.

## DAFTAR PUSTAKA

- [1] Silvia Estefina Subitmele, "Profil PT PGN, Perusahaan yang Bergerak di Bidang Transmisi dan Distribusi Gas Bumi," *Liputan* 6.
- [2] Y. Suh, "Discovering customer segments through interaction behaviors for home appliance business," *J Big Data*, vol. 12, no. 1, p. 57, Mar. 2025, doi: 10.1186/s40537-025-01111-y.
- [3] A. Karami and Y. Widharto, "PERANCANGAN BUSINESS INTELLIGENCE DAN SEGMENTASI PELANGGAN MENGGUNAKAN K MEANS CLUSTERING BERDAARKAN RFM MODEL," *Industrial Engineering Online Journal*, vol. 12, no. 1, Jan. 2023.
- [4] Y. Yuan and Z. Wang, "A Data-Driven Customer Segmentation Strategy Based on Contribution to System Peak Demand," *ArXiv*, Oct. 2018.
- [5] N. M. Br. Tarigan, S. E. Br. Tarigan, and A. P. Simatupang, "Implementation of Data Mining in Grouping Data of the Poor Using the K-Means Method," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 5, no. 2, pp. 599–611, Aug. 2023, doi: 10.47709/cnahpc.v5i2.2625.
- [6] Eka Angga Laksana and Marchel Maulana Fahrezi, "Costumer Segmentation and Analysis Based on Gaussian Mixture Model Alghorithm," *Proceedings of the Widyatama International Conference on Engineering 2024*, pp. 67–74, 2024.
- [7] Danu Kuncoro Aji and Chaulina Alfianti Oktavia, "SISTEM INFORMASI GEOGRAFIS PEMETAAN ANALISIS DAERAH RAWAN KRIMINALITAS MENGGUNAKAN METODE JENKS NATURAL BREAKS BERBASIS WEB KOTA MALANG," *ELANG: Journal of Interdisciplinary Research*, pp. 13–23, Jun. 2023.
- [8] J. Wang, "Analysis of Shared Bicycle Usage based on K-Means and GMM Clustering Algorithm," in *2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, IEEE, Oct. 2021, pp. 92–96. doi: 10.1109/AINIT54228.2021.00028.
- [9] Widiastari and Afrialita, "Analisa Datamining Dengan Metode Klasifikasi C4.5 Sebagai Faktor Penyebab Tanah Longsor," *Journal of Computer System and Informatics (JoSYC)*, no. 3, p. 55, 2021.
- [10] Horainim Sibarani, Solikhun, Widodo Saputra, Indra Gunawan, and Zulaini Masruro Nasution, "PENERAPAN METODE K-MEANS UNTUK PENGELOMPOKKAN KABUPATEN/KOTA DI PROVINSI SUMATERA UTARA BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN MANUSIA," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, pp. 154–161, Feb. 2022.
- [11] M. Shutaywi and N. N. Kachouie, "Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering," *Entropy*, vol. 23, no. 6, p. 759, Jun. 2021, doi: 10.3390/e23060759.
- [12] M. Saleh, "Evaluation of Jenks Natural Breaks Clustering Algorithm for Changeoint Identification in Streaming Sensor Data," *IEEE Sens Lett*, vol. 8, no. 10, pp. 1–4, Oct. 2024, doi: 10.1109/LSSENS.2024.3456292.
- [13] I. P. Putra and A. Fadhillah, "Perbandingan Metode K-Means dan Hierarchical Clustering dalam Pengelompokan Data Penduduk Miskin di Kabupaten Cianjur," *LANCAH: Jurnal Inovasi dan Tren*, vol. 3, no. 1, Apr. 2025, doi: 10.35870/ljit.v3i1.4028.
- [14] D. Anggreani, N. Nurmisba, D. Setiawan, and L. Lukman, "Optimization of K-Means Clustering Method by Using Elbow Method in Predicting Blood Requirement of Pelamonia Hospital Makassar," *Internet of Things and Artificial Intelligence Journal*, vol. 4, no. 3, pp. 541–550, Aug. 2024, doi: 10.31763/iota.v4i3.755.
- [15] M. Butwall, P. Ranka, and S. Shah, "Python in Field of Data Science: A Review," *Int J Comput Appl*, vol. 178, no. 49, pp. 20–24, Sep. 2019, doi: 10.5120/ijca2019919404.