

ANALISIS *CLUSTER* TERHADAP TINGKAT PENCEMARAN UDARA PADA SEKTOR INDUSTRI DI SULAWESI TENGGARA

Sarjum Bao¹⁾

¹⁾Program Studi Matematika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Halu Oleo, Kendari, Indonesia
Email: sarjumbao22@gmail.com

Arman^{1,a)}, La Gubu^{1,b)}, Wayan Somayasa^{1,c)}, Bahriddin^{2,d)} dan Agusrawati^{2,e)}

¹⁾Program Studi Matematika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Halu Oleo, Kendari, Indonesia

²⁾Program Studi Statistika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Halu Oleo, Kendari, Indonesia

Email:^{a)}arman.mtmk@uho.ac.id, ^{b)}la.gubu@uho.ac.id, ^{c)}wayan.somayasa@uho.ac.id,

^{d)}bahriddin.abapihi@uho.ac.id dan ^{e)}agusrawati@uho.ac.id

ABSTRAK

Pencemaran udara semakin hari dirasakan semakin meningkat, terutama di daerah dengan kepadatan lalu lintas yang cukup tinggi, dan di lokasi perindustrian yang kurang memperhatikan dampak buruk pada lingkungan. Tujuan dari penelitian ini adalah untuk mengetahui penerapan cluster dengan menggunakan metode hierarki dan metode non hierarki pada industri yang ada di Sulawesi Tenggara sehingga dapat memberikan pemahaman tentang cara mengatasi polusi udara berlebih. Penyelesaiannya diawali dengan melakukan analisis deskriptif yang dimana untuk melihat nilai minimal (*min*), nilai rata-rata (*mean*), dan nilai maksimal (*maks*). Selanjutnya melakukan deteksi *outlier* dan setelah itu melakukan uji asumsi *cluster* dengan menggunakan uji sampel representatif dan uji multikolinearitas agar data dapat digunakan untuk proses pengelompokkan. Normalisasi *min-max* dilakukan sebelum melakukan proses klusterisasi. Menerapkan analisis *cluster* dengan menentukan ukuran jarak *euclidean*, membuat *cluster* dengan menggunakan metode hierarki dan metode *non* hierarki serta menentukan jarak ketidakmiripan antara sebuah objek. Menginterpretasikan hasil *cluster* yang diperoleh dua metode dengan data pencemaran udara pada sektor industri dari Dinas Lingkungan Hidup (DLH) Sulawesi Tenggara.

Kata kunci : Pencemaran Udara pada Sektor Industri, Analisis Cluster, Analisis Deskriptif, Deteksi Outlier, Dinas Lingkungan Hidup Sulawesi Tenggara.

ABSTRACT

Air pollution is increasingly being felt, especially in areas with high traffic density, and in industrial locations that pay little attention to the negative impact on the environment. The purpose of this study is to determine the application of clusters using hierarchical and non-hierarchical methods in industries in Southeast Sulawesi so as to provide an understanding of how to deal with excess air pollution. The solution begins with conducting a descriptive analysis in which to see the minimum value (*min*), the average value (*mean*), and the maximum value (*max*). Then perform outlier detection and after that carry out a cluster assumption test using a representative sample test and multicollinearity test so that the data can be used for the grouping process. Min-max normalization is carried out before carrying out the clustering process. Applying cluster analysis by determining the size of the Euclidean distance, creating clusters using hierarchical and non-hierarchical methods and determining the dissimilarity distance between objects. Interpret the cluster results obtained by the two methods with air pollution data in the industrial sector from the Southeast Sulawesi Environmental Service (DLH).

Keywords: Air Pollution in the Industrial Sector, Cluster Analysis, Deskriptif Analysis, Outlier Detection, Southeast Sulawesi Environmental Service.

1. Pendahuluan

Industri menempati posisi yang cukup sentral dalam siklus perekonomian masyarakat modern sekaligus penggerak paling dasar dalam peningkatan kemakmuran, terutama pada negara-negara maju. Sedangkan pada negara berkembang, industri sangat penting dalam perluasan landasan pembangunan dan

memenuhi kebutuhan masyarakat yang kian meningkat. Sementara itu, kondisi lingkungan

khususnya udara sangat dipengaruhi oleh faktor-faktor yang salah satunya yaitu faktor industri.

Pembangunan pada sektor industri membawa dampak yang signifikan bagi kehidupan manusia, baik dampak positif maupun negatif. Salah satu dampak

negatif dari pembangunan pada sektor industri yaitu pencemaran udara. Pencemaran udara semakin hari dirasakan semakin meningkat, terutama di daerah dengan kepadatan lalu lintas yang cukup tinggi, dan di lokasi perindustrian yang kurang memperhatikan dampak buruk pada lingkungan. Untuk mewaspadai akibat dari masalah polusi, pertama-tama kita harus mengetahui kapan saja kondusi polusi udara berbahaya bagi kesehatan manusia. Salah satunya yaitu dengan memprediksi dengan melihat pola dari polusi udara yang telah terjadi beberapa tahun terakhir sehingga masyarakat lebih waspada dan dapat mencegah akibat negatif dari polusi udara [7].

Dalam analisis statistik, kriteria polutan tersebut dinyatakan sebagai variabel. Semakin banyak kriteria polutan, akan semakin sulit analisis statistika yang akan dilakukan. Analisis multivariat merupakan salah satu metode statistika yang cocok untuk meringkas data dengan peubah banyak. Salah satunya yaitu analisis *cluster*. Analisis *cluster* mempunyai tujuan yaitu untuk mengelompokkan objek-objek berdasarkan karakteristik diantara objek-objek itu, sehingga dapat diketahui ciri khas dari tiap-tiap kelompok. Beberapa objek yang dapat dikelompokkan dengan analisis *cluster*, diantaranya yaitu produk (barang dan jasa), benda, dan manusia [7].

Analisis *cluster* merupakan suatu teknik statistik multivariat untuk mengidentifikasi satu *cluster* objek yang mempunyai kesamaan karakteristik tertentu yang dapat dipisahkan dengan *cluster* objek yang lain, sehingga perbedaan setiap *cluster* dapat dilihat dengan jelas. Analisis *cluster* mengelompokkan objek seperti responden, produk, tanaman, hewan dan lain-lain sehingga masing-masing objek yang sangat mirip dengan objek lainnya akan bergabung kedalam satu *cluster* dengan kriteria seleksi yang telah ditentukan sebelumnya. Hasil analisis *cluster* seharusnya menunjukkan kehomogenan objek-objek yang tinggi dalam *cluster* dan keheterogenan yang tinggi antar *cluster*. Tujuan pokok dari analisis *cluster* adalah untuk mengelompokkan objek pengamatan menjadi beberapa kelompok yang lebih sederhana berdasarkan tingkat kehomogenan objek pengamatan [2].

Berdasarkan uraian diatas, maka latar belakang dari penelitian ini adalah untuk mengetahui hasil klasterisasi terhadap tingkat pencemaran udara pada sektor industri di Sulawesi Tenggara.

Pada bagian dua membahas tentang analisis multivariat, analisis *cluster*, konsep dasar analisis *cluster*, algoritma *K-Means*, normalisasi *Min-Max*, ukuran kedekatan jarak, dan asumsi analisis *cluster*. Pada bagian tiga dijelaskan mengenai metode penelitian yang akan dilakukan pada penelitian ini. Pada bagian empat menjelaskan tentang hasil penelitian dan pembahasan dari penelitian yang telah

dilakukan. Pada bagian lima membahas tentang kesimpulan dan saran.

2. Tinjauan Pustaka

2.1 Analisis Multivariat

Analisis multivariat merupakan analisis yang meneliti tentang masalah yang bersifat multidimensional dan menggunakan tiga atau lebih variabel. Berikut pembagian dua kategori utama Analisis Multivariat [7].

1. Dependence Method

Analisis ketergantungan digunakan apabila tujuan dari analisis adalah menjelaskan atau memprediksi variabel terikat berdasarkan dua atau lebih variabel bebas. Metode ini terdiri atas empat jenis yaitu: Analisis Regresi Berganda, Analisis Diskriminan Berganda, Analisis Multivariat Varians, Analisis Korelasi Kanonikal.

2. Interdependence Method

Merupakan metode yang digunakan untuk menjelaskan seperangkat variabel atau pengelompokkan berdasarkan variabel-variabel tertentu. Metode ini dikelompokkan menjadi tiga yaitu Analisis Faktor, Analisis *Cluster*, Skala Multidimensional

2.2 Analisis Cluster

Analisis *cluster* adalah suatu alat untuk mengelompokkan sejumlah n objek berdasarkan p variabel yang secara relatif mempunyai kesamaan karakteristik diantara objek-objek tersebut, sehingga keragaman dalam suatu kelompok tersebut lebih kecil dibandingkan dengan keragaman antar kelompok. Objeknya dapat berupa barang, jasa, hewan, manusia (responden, konsumen, atau yang lain). Objek tersebut akan diklarifikasikan dalam satu atau lebih *cluster* (kelompok) sehingga objek-objek yang berada dalam satu *cluster* akan mempunyai kemiripan atau kesamaan karakter [7].

2.2.1 Metode Hierarki

Metode Hierarki digunakan untuk mengelompokkan objek secara terstruktur berdasarkan kemiripan sifatnya dan *cluster* yang diinginkan belum diketahui banyaknya [8]. Berikut penjelasan mengenai beberapa pautan dalam metode pengklasteran hierarki:

1) Pautan Tunggal (*Single Linkage*)

Jarak dua *cluster* diukur dengan jarak terdekat antara sebuah objek dalam *cluster* yang satu dengan yang lain.

$$D(uv)w = \min(d_{uw}, d_{vw})$$

2) Pautan Lengkap (*Complete Linkage*)

Jarak dua *cluster* diukur dengan jarak terjauh antara sebuah objek dalam *cluster* yang satu dengan yang lain

$$D(uv)w = \max(d_{uw}, d_{vw})$$

3) Pautan Centroid (*Centroid Linkage*)

Jarak antara dua buah *cluster* diukur sebagai jarak *Euclidian* antara kedua rata-rata (*centroid*) *cluster*.

$$D_{(ij)k} = \frac{n_i}{n_i + n_j} d_{ik} + \frac{n_j}{n_i + n_j} d_{jk} - \frac{n_i n_j}{n_i + n_j} d_{ij}$$

4) Pautan Rataan (*Average Linkage*)

Jarak antara dua *cluster* diukur dengan jarak rata-rata antara sebuah objek dalam *cluster* yang satu dengan yang lain

$$D_{(uv)w} = \frac{n_u}{n_u + n_v} d_{uw} + \frac{n_v}{n_u + n_v} d_{vw}$$

5) Metode Ward (*Ward Method*)

Jarak antar dua *cluster* adalah total jumlah kuadrat dua *cluster* pada masing-masing variabel. Metode ini berbeda dengan metode lainnya karena menggunakan pendekatan analisis varians untuk menghitung jarak antar *cluster* atau metode ini meminimumkan jumlah kuadrat (ESS)

$$ESS = \sum_{j=1}^n x_j^2 - \frac{1}{n} \left(\sum_j x_j \right)^2$$

2.2.2 Metode Non Hierarki

Metode non hierarki digunakan untuk mengelompokkan objek dimana banyak *cluster* yang akan dibentuk dapat ditentukan terlebih dahulu sebagai bagian dari prosedur penggerombolan. Kemudian metode ini dapat diterapkan pada data yang lebih besar dibandingkan dengan metode hierarki

2.3 Konsep Dasar Analisis Cluster

2.3.1 Menentukan Ukuran Ketidakmiripan Antara Kedua Objek

Salah satu yang bisa menjadi ukuran ketidakmiripan adalah fungsi jarak antara objek *a* dan *b*, yang biasa dinotasikan dengan $d(a, b)$. Adapun sifat-sifat ukuran ketidakmiripan adalah:

- $d(a, b) \geq 0$
- $d(a, b) = 0$
- $d(a, b) = d(b, a)$
- (a, b) meningkat seiring semakin tidak miripnya kedua objek *a* dan *b*
- $d(a, c) \leq d(a, b) + d(b, c)$

Jarak yang digunakan adalah jarak *Euclidian*. Ukuran jarak atau ketidakmiripan antar objek ke-*i* dengan objek ke-*h* dapat disimbolkan dengan d_{ih} . Adapun nilai d_{ih} diperoleh melalui perhitungan jarak kuadrat sebagai berikut:

$$d_{ih} = \sum_{j=1}^m (x_{ij} - x_{hj})^2$$

Dimana:

d_{ih} = jarak kuadrat *Euclidian* antara objek ke-*i* dengan objek ke-*h*

m = banyaknya variabel yang digunakan

x_{ij} = nilai dari objek ke-*i* pada variabel ke-*j*

x_{hj} = nilai dari objek ke-*h* pada variabel ke-*j*

2.3.2 Membuat Cluster

Proses pembuatan *cluster* dapat dilakukan dengan 2 tahap, yaitu:

1. Metode Hierarki

Pada metode ini, dimulai dengan mengelompokkan data yang mempunyai kesamaan

yang paling dekat. Kemudian diteruskan ke objek lain yang mempunyai kedekatan yang kedua. Demikian seterusnya sehingga terbentuk semacam “pohon”.

2. Metode Non Hierarki

Metode ini dimulai dengan proses penentuan jumlah gerombol terlebih dahulu, dan yang digunakan adalah metode K-Means. Metode K-Means digunakan sebagai alternatif metode cluster untuk data dengan ukuran yang besar karena memiliki kecepatan yang lebih tinggi dibandingkan metode hierarki

2.3.3 Melakukan Interpartasi terhadap cluster yang terbentuk

2.4 Algoritma K-Means

K-Means termasuk dalam *partitioning clustering* yaitu setiap data harus masuk dalam *cluster* tertentu untuk memungkinkan bagi setiap datayang termasuk dalam *cluster* tertentu pada satu tahapan proses, pada tahapan berikutnya berpindah ke *cluster* yang lain. Berikut adalah langkah-langkah algoritma *K-Means* :

- Tentukan *k* sebagai jumlah *cluster* yang dibentuk
- Tentukan pusat (*centroid*) *cluster* awal.
- Perhitungan jarak dengan pusat *cluster*
- Mengalokasikan masing-masing data ke *centroid*/rata-rata terdekat
- Mengelompokkan data berdasarkan jarak terpendeknya antara data dengan *centroid* menjadi sebuah kelompok *cluster*.

2.5 Normalisasi Min-Max

Normalisasi min-max merupakan metode normalisasi yang mengubah nilai pada data secara linear, yaitu mengubah nilai maksimum menjadi 1 dan nilai minimum menjadi 0.

Persamaan dari normalisasi *min-max* yaitu sebagai berikut:

$$d' = \frac{d - \min(p)}{\max(p) - \min(p)}$$

Dimana:

d' = nilai hasil normalisasi

d = nilai sebelum normalisasi

p = keseluruhan nilai pada suatu atribut

$\max(p)$ = nilai maksimum pada suatu atribut

$\min(p)$ = nilai pada suatu atribut

2.6 Ukuran Kedekatan Jarak

Ukuran kedekatan biasanya digunakan dalam pengelompokkan untuk membentuk *cluster-cluster* individu berdasarkan kedekatan antar masing-masing individu sehingga keragaman antar individu dalam

cluster lebih kecil daripada keragaman antar *cluster*. Ukuran kedekatan dapat diperoleh dengan:

- Penilaian pasangan-pasangan objek dengan pertimbangan pada kesamaan atau ketidaksamaan secara langsung. Data yang diperoleh disebut kemiripan langsung (*direct similarities*)
- Penilaian terhadap peubah-peubah yang telah diteapkan. Data yang diperoleh disebut dengan kemiripan turunan (*derived similarities*).

Setelah data dikumpulkan, ukuran kedekatan dapat dihitung dengan menggunakan ukuran jarak. Sebagai contoh suatu variabel x_1, x_2, \dots, x_n dengan banyak sampel n masing-masing s_1, s_2, \dots, s_n dari variabel x_1 dan x_n . Dari masing-masing variabel x_1 dihubungkan dengan sampel s_n . Sehingga dapat dimodelkan x_{ij} adalah:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}$$

(2.11)

Dengan: $i = 1, 2, \dots, m$
 $j = 1, 2, \dots, n$

2.6.1 Jarak Euclid (Euclidean Distance)

Jarak *Euclid* merupakan tipe pengukuran jarak dalam analisis *cluster* yang paling umum digunakan untuk mengukur jarak dari objek data ke pusat *cluster*.

Dalam bentuk umum, jarak *Euclid* (d) dapat diperoleh dengan:

$$x = (x_1, x_2, \dots, x_n)$$

$$y = (y_1, y_2, \dots, y_n)$$

$$d(x, y) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \dots + (y_n - x_n)^2}$$

Atau secara umum:

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}$$

(2.14)

Dimana:

d_{ij} = jarak *Euclid* objek data ke- i dan objek data ke- j

m = banyaknya peubah/parameter yang digunakan

x_{ik} = objek data ke- i pada peubah ke- k

x_{jk} = objek data ke- j pada peubah ke- k

2.6.2 Jarak Mahalanobis

Jarak *Mahalanobis* merupakan generalisasi dari jarak kuadrat *Euclid* yang distandarisasi. Selain itu, jarak *Mahalanobis* tidak hanya dapat mengatasi masalah perbedaan skala dalam data, tetapi juga mempertimbangkan pengaruh korelasi antar peubah. Pada saat peubah-peubah tidak saling berkorelasi atau (r) = 0, jarak *Mahalanobis* sama dengan jarak *Euclid* yang distandarisasi. Jarak *Euclid* merupakan kasus khusus dari jarak *Mahalanobis*.

Jarak mahalanobis didefinisikan:

$$d_{(i,j)} = [(\bar{x}_i - \bar{x}_j)s^{-1}(\bar{x}_i - \bar{x}_j)]^{\frac{1}{2}}$$

(2.15)

Dengan:

$d_{(i,j)}$: jarak antara objek i ke objek j

\bar{x}_i : nilai tengah pada *cluster* ke- i

\bar{x}_j : nilai tengah pada *cluster* ke- j

s^{-1} : matriks varians dan kovarians gabungan

2.7 Asumsi Analisis Cluster

2.7.1 Sampel Representatif

Sampel representatif adalah sampel yang diambil dapat dikatakan mempresentasikan atau mewakili populasi yang ada. Metode Kaiser-Mayer-Olkin (KMO) digunakan untuk mengukur kecukupan *sampling* secara menyeluruh dan mengukur kecukupan *sampling* untuk setiap indikator. Sampel dikatakan representatif apabila nilai KMO berkisar antara 0,5 sampai 1

$$KMO = \frac{\sum_{j=1}^p \sum_{k=1}^p r_{x_j x_k}^2}{\sum_{j=1}^p \sum_{k=1}^p r_{x_j x_k}^2 + \sum_{j=1}^p \sum_{k=1}^p \rho_{x_j x_k x_l}^2}$$

(2.16)

Dengan:

$$r_{x_j x_k}^2 = \frac{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n}} \sqrt{\frac{\sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}{n}}}$$

$$\rho_{x_j x_k x_l}^2 = \frac{r_{x_j x_k} r_{x_j x_l} r_{x_k x_l}}{\sqrt{(1 - r_{x_j x_l}^2)} \sqrt{1 - r_{x_k x_l}^2}}$$

$r_{x_j x_k}$: Korelasi antar variabel x_j dan x_k

\bar{X}_k : Rata-rata variabel X_k

\bar{X}_j : Rata-rata variabel X_j

$\rho_{x_j x_k x_l}^2$: Korelasi parsial antara variabel X_j

dan X_k dengan menjaga agar X_l konstan.

2.7.2 Uji Multikolinearitas

Multikolinearitas adalah hubungan linear yang ada di antara variabel independen. Multikolinearitas dapat dilihat dari nilai Variance Inflation Factor (VIF). Rumus untuk menghitung VIF yaitu sebagai berikut:

$$VIF_j = \frac{1}{tol}$$

(2.17)

$$tol = (1 - R^2)$$

Dengan:

R^2 : kuadrat nilai korelasi antar variabel bebas

Jika nilai VIF kurang dari angka 10 dan nilai toleransi lebih besar dari 0,1 maka dapat disimpulkan tidak terjadi multikolinearitas.

3. Metode

Metode yang digunakan pada penelitian ini menggunakan metode kepustakaan atau studi literatur dengan langkah-langkah sebagai berikut :

- Mengidentifikasi masalah
- Pengambilan data pencemaran udara pada sektor industri dari Dinas Lingkungan Hidup (DLH) Provinsi Sulawesi Tenggara
- Menentukan Variabel dan Objek
- Menginput data penelitian
- Melakukan analisis deskriptif

6. Melakukan uji asumsi *cluster*
7. Melakukan Normalisasi pada data dengan Normalisasi *Min-Max*
8. Menerapkan analisis *cluster* pada data penelitian
9. Interpretasi hasil *cluster* yang diperoleh dari dua metode pada langkah kedua
10. Menarik kesimpulan

4. Hasil dan Pembahasan

4.1 Analisis Deskriptif

Data yang digunakan merupakan data pencemaran udara pada sektor industri yang diperoleh dari Dinas Lingkungan Hidup (DLH) Provinsi Sulawesi Tenggara tahun 2022

Tabel Error! No text of specified style in document.1
 Beban Pencemaran Udara pada 8 Sektor Industri di Sultra

No	Jenis Industri	TSP	CO	NH3	NO2	SO2	Satuan
1	Antam	24,34	363,71	0,75	2746,1	1823,08	$\mu\text{g}/\text{nm}^3$
2	PLTD Wangi-wangi	46,20	509	0,26	883	40,32	$\mu\text{g}/\text{nm}^3$
3	PT Merbaujaya	53,32	562	0,15	2183	19,01	$\mu\text{g}/\text{nm}^3$
4	PT DSSP Power	410,60	3098,08	0,87	61,30	276,40	$\mu\text{g}/\text{nm}^3$
5	Pertamina Raha	66,10	1454	0	9,70	42,50	$\mu\text{g}/\text{nm}^3$
6	Pertamina Kendari	135,50	4695	0	21,50	47,60	$\mu\text{g}/\text{nm}^3$
7	Pertamina Kolaka	79,20	1007	0	9,68	41	$\mu\text{g}/\text{nm}^3$
8	PLN Kolaka Utara	47,10	452	0,97	323	12,40	$\mu\text{g}/\text{nm}^3$

Tabel 4.2 Statistik Deskriptif

Variabel	Indeks Polutan total	rata-rata	Minimal	Maksimal	Satuan
TSP/Debu	738,50	123	46,20	410,60	$\mu\text{g}/\text{nm}^3$
CO (Carbon Monoksida)	11.777,08	1.472,14	363,71	4695	$\mu\text{g}/\text{nm}^3$
Amonia (NH ₃)	3,01	0,37	0	0,97	$\mu\text{g}/\text{nm}^3$
NO ₂ (Nitrogen Dioksida)	3491,18	436,39	9,68	2746,14	$\mu\text{g}/\text{nm}^3$
SO ₂ (Sulfur Dioksida)	419,90	52,48	12,40	1823,08	$\mu\text{g}/\text{nm}^3$

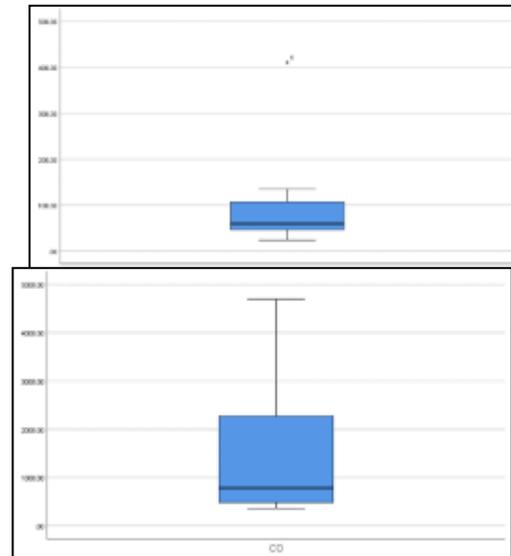
Dari Tabel Error! No text of specified style in document.1 dan Tabel 4.2, dapat dilihat pula bahwa data yang digunakan memiliki jangkauan yang berbeda. Oleh karena itu, Sebelum melakukan proses analisis *cluster*, terlebih dahulu dilakukan pendeteksian terhadap *outlier* untuk mengetahui apakah data tersebut layak di gunakan dalam proses analisis *cluster* atau tidak.

4.2 Deteksi Outlier

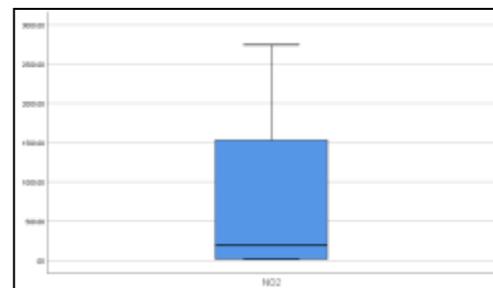
Sebelum melakukan uji asumsi analisis *cluster*, data perlu di lakukan deteksi apakah terdapat *outlier* atau tidak. *Outlier* berarti bahwa terdapat data yang

memiliki perbedaan yang signifikan dari pada data lainnya.

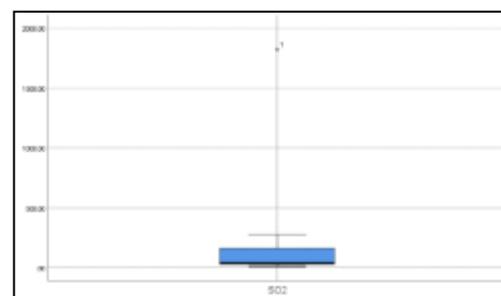
Gambar 4.1 *Boxplot* TSP
 (Sumber: Output SPSS)



Gambar 4.3 *Boxplot* NH₃
 (Sumber: Output SPSS)



Gambar 4.4 *Boxplot* NO₂
 (Sumber: Output SPSS)



Gambar 4.5 *Boxplot* SO₂
 (Sumber: Output SPSS)

Berdasarkan hasil pendeteksian *outlier* tersebut, peneliti memutuskan untuk tidak mengikutsertakan data *outlier* dalam analisis *cluster*. Hal ini disebabkan karena jika data *outlier*

diikutsertakan dalam analisis, dapat menyebabkan hasil *cluster* yang kurang representatif. Oleh karena itu, data *outlier* tersebut tidak diikutsertakan dalam analisis *cluster*

4.3 Uji Asumsi Cluster

4.3.1 Uji Sampel Representatif

Pengujian sampel representatif akan dilakukan dengan menggunakan *software* SPSS 26. Berikut merupakan hasil pengujian sampel representatif dengan uji Kaiser-Meyer-Olkin (KMO) menggunakan *software* SPSS 26.

Tabel 4.3 Output Hasil Uji KMO
 (Sumber: Output SPSS)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.51
		9
Bartlett's Test of Sphericity	Approx. Chi-Square	6.3
		01
	df	10
	Sig.	.78
		9

Berdasarkan hasil *output* perhitungan yang dilakukan dengan menggunakan *software* SPSS 26, maka sesuai penjelasan pada subbab 2.6.1, sampel sudah layak dianalisis.

4.3.2 Uji Multikolinearitas

Tabel 4.4 Hasil Output Uji Multikolinearitas
 (Sumber: Output SPSS)

Coefficients ^a			
Model		Collinearity Statistics	
		Tolerance	VIF
1	TSP	.387	2.586
	CO	.638	1.566
	NH3	.452	2.212
	NO2	.735	1.361

a. Dependent Variable: SO2

Tabel 4.4 merupakan *output* dari perhitungan uji multikolinearitas dengan menggunakan *Software* SPSS 26. Suatu data dikatakan terbebas dari multikolinearitas apabila nilai *VIF* < 10. Pada Tabel 4.4, dapat diperhatikan bahwa nilai *VIF* dari keempat variabel adalah < 10. Hal tersebut memiliki arti bahwa keempat variabel dapat digunakan untuk langkah selanjutnya yaitu proses *clustering*.

4.4 Normalisasi Min-Max

Pada kasus ini, normalisasi yang digunakan yaitu normalisasi *min-max* dengan menggunakan Persamaan 2.10, sehingga diperoleh:

Untuk objek jenis industri Antam:

$$d' = \frac{d - \min(p)}{\max(p) - \min(p)}$$

$$d'_{TSP} = \frac{24,34 - 0}{135,5 - 0} = 0,17$$

$$d'_{CO} = \frac{363,71 - 363,71}{4695 - 363,71} = 0$$

$$d'_{NH_3} = \frac{0,75 - 0}{0,972 - 0} = 0,77$$

NO	Jenis Industri	TSP	CO	NH ₃	NO ₂	SO ₂	Satuan
1	Antam	0,17	0	0,77	1	0	µg/nm ³
2	PLTD	0,34	0,03	0,26	0,31	0,14	µg/nm ³
3	PT Merbaujaya	0,39	0,04	0,15	0,79	0,06	µg/nm ³
4	PT DSSP Power	0	0,63	0,9	0,01	1	µg/nm ³
5	Pertamina Raha	0,48	0,25	0	0	0,15	µg/nm ³
6	Pertamina Kendari	1	1	0	0	0,17	µg/nm ³
7	Pertamina Kolaka	0,58	0,14	0	0	0,14	µg/nm ³
8	PLN Kolaka Utara	0,34	0,02	1	0,11	0,04	µg/nm ³

Tabel 4.5 Normalisasi *Min-Max*

$$d'_{NO_2} = \frac{2746,14 - 9,68}{2746,14 - 9,68} = 1$$

$$d'_{SO_2} = \frac{0 - 0}{276,4 - 0} = 0$$

Dan seterusnya dengan cara yang sama, kemudian akan didapatkan hasil perhitungan normalisasi yang disajikan dalam Tabel 4.5 berikut ini:

4.5 Jarak Euclidean

Data yang telah di uji menggunakan uji asumsi *cluster* dan telah dinormalisasi kemudian akan dianalisis menggunakan analisis *cluster*. Pada penelitian ini, metode analisis *cluster* yang digunakan yaitu metode hierarki dan metode non hierarki, dimana metode hierarki yaitu dengan menggunakan metode *average linkage* dan *complete linkage*, sementara metode non hierarki dengan menggunakan metode *K-Means*. Kemudian akan dibandingkan menggunakan lima variabel (polutan), tujuannya adalah untuk menemukan dua industri yang paling mirip di antara kedelapannya.

Pertamina Raha dan Pertamina Kolaka tergabung dalam 1 *cluster*. Semakin rendah atau kecil nilai suatu jarak, maka semakin tinggi tingkat kemiripan karakteristik antar data. Hasil perhitungan

jarak *Euclidean* tersebut juga bisa di tampilkan dalam bentuk tabel proximity matriks seperti pada Tabel berikut ini:

Tabel 4.6 Tabel Proximity Matriks Jarak *Euclidean*

D_{Euclid}	1	2	3	4	5	6	7	8
1	0	1	0,68	1,55	1,33	1,81	1,34	0,93
2		0	0,49	1,3	0,49	1,24	0,49	0,76
3			0	1,58	0,84	1,39	0,84	1,08
4				0	1,38	1,62	1,45	1,19
5					0	0,9	0,14	1,04
6						0	0,94	1,55
7							0	1,04
8								0

4.6 Proses Analisis Cluster

4.6.1 Metode *Complete Linkage*

Pada awal proses terdapat 8 *cluster*. Tahap pertama yang dilakukan adalah menentukan jarak yang terdekat antara dua objek dari sekian banyaknya kombinasi jarak dari kedelapan objek yang ada. jarak antara antara Pertamina Raha dan Pertamina Kolaka memiliki kemiripan jarak atau memiliki jarak yang paling dekat yaitu 0,14. Oleh karena itu, Pertamina Raha dan Pertamina Kolaka tergabung dalam 1 *cluster*. Lalu setelah digabungkan maka sekarang tersisa 7 objek.

Kemudian dilakukan perbaikan matriks jarak menggunakan metode *Complete Linkage* yang akan dihitung manual, sehingga terjadi perubahan jarak yang melibatkan *cluster* baru (*cluster* yang anggotanya Pertamina Raha dan Pertamina Kolaka) sehingga diperoleh:

$$D(5,7)1 = \max(d_{51}, d_{71}) = \max(1,33, 1,34) = 1,34$$

$$D(5,7)2 = \max(d_{52}, d_{72}) = \max(0,49, 0,5) = 0,5$$

$$D(5,7)3 = \max(d_{53}, d_{73}) = \max(0,85, 0,84) = 0,845$$

$$D(5,7)4 = \max(d_{54}, d_{74}) = \max(1,38, 1,45) = 1,45$$

$$D(5,7)6 = \max(d_{56}, d_{76}) = \max(0,9, 0,94) = 0,94$$

$$D(5,7)8 = \max(d_{58}, d_{78}) = \max(1,05, 1,04) = 1,05$$

Dari penyelesaian menggunakan Persamaan 2.2 di atas, maka hasil *cluster* tersebut dapat di sajikan dalam bentuk Tabel berikut ini:

Tabel 4.8 Proximity Matriks Metode *Complete Linkage*

D_{Euclid}	(5,7)	1	2	3	4	6	8
(5,7)	0	1,34	0,5	0,84	1,45	0,94	1,05
1		0	1	0,68	1,55	1,81	0,93
2			0	0,49	1,3	1,24	0,76
3				0	1,58	1,39	1,08
4					0	1,62	1,19
6						0	1,55
8							0

Dan seterusnya dilakukan perhitungan jarak dari data-data yang telah di klusterisasikan dengan menggunakan metode *Average Linkage*. Proses pengklasteran dilakukan hingga tahap 5, kemudian akan didapatkan hasil akhir pengklasteran yang disajikan dalam Tabel berikut ini:

Tabel 4.7 Proximity Matriks *Complete Linkage final step cluster*

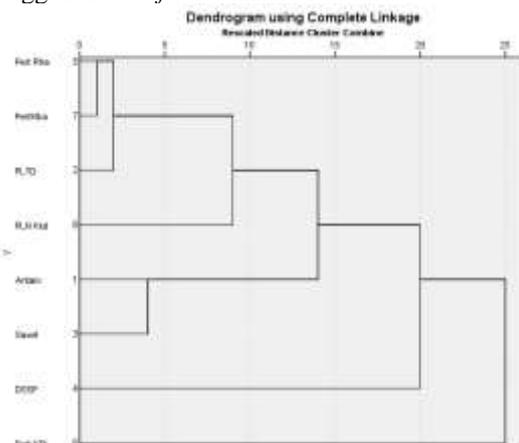
D_{Euclid}	(1,2,3,5,7,8)	4	6
(1,2,3,5,7,8)	0	1,59	1,81
4		0	1,62
6			0

Dari Tabel 4.36 di atas, dapat dilihat bahwa jarak antara objek (1,2,3,5,7,8) dengan objek 4 memiliki jarak terdekat, yakni 1.587. Maka objek-objek tersebut di kelompokkan menjadi 1 *cluster*. Setelah objek-objek tersebut sudah dikelompokkan, maka sekarang tersisa 2 *cluster*. Sehingga terdapat 2 *cluster* yang terbentuk pada tahap terakhir dari metode *Complete Linkage* ini yakni tahap 5, dimana *cluster-cluster* tersebut terdiri dari:

Cluster 1: Pertamina Kendari (6)

Cluster 2: Industri Antam, PLTD Wangi-wangi, PT Merbaujaya Indahraya, PT DSSP Power Kendari, Pertamina Raha, Pertamina Kolaka, PLN Kolaka Utara (1,2,3,4,5,7,8).

Proses pengelompokkan dengan metode *Complete Linkage* di atas juga dapat diinterpretasikan kedalam dendrogram. Dendrogram dapat dibaca dari kiri ke kanan. Garis vertikal pada dendrogram menunjukkan jarak *cluster* yang digabungkan. Sedangkan garis horizontal menunjukkan *cluster* yang digabung bersama. Berikut ini merupakan hasil dendrogram dengan metode *Complete Linkage* dengan menggunakan *Software IBM SPSS 26*.



Gambar 4.6 Dendrogram Metode *Complete Linkage*

4.6.2 Metode *Average Linkage*

Jarak antara antara Pertamina Raha dan Pertamina Kolaka memiliki kemiripan jarak atau memiliki jarak yang paling dekat yaitu 0,141. Oleh

karena itu, Pertamina Raha dan Pertamina Kolaka bergabung dalam 1 cluster.

Selanjutnya, dilakukan perbaikan matriks jarak dengan cara menghitung nilai rata-rata dari setiap variabel pada tiap-tiap objek industri yang dipasangkan yang akan dihitung manual sehingga terjadi perubahan jarak yang melibatkan cluster baru (cluster yang anggotanya Pertamina Raha dan Pertamina Kolaka). Sehingga diperoleh:

$$D_{(uv)w} = \frac{n_u}{n_u + n_v} d_{(uw)} + \frac{n_v}{n_u + n_v} d_{(vw)}$$

$$D_{(5,7)1} = \frac{n_5}{n_5 + n_7} d_{(5,1)} + \frac{n_7}{n_5 + n_7} d_{(7,1)}$$

$$D_{(5,7)1} = \frac{1}{2} 1,33 + \frac{1}{2} 1,34$$

$$D_{(5,7)1} = \frac{2,67}{2} = 1,33$$

$$D_{(5,7)2} = \frac{n_5}{n_5 + n_7} d_{(5,2)} + \frac{n_7}{n_5 + n_7} d_{(7,2)}$$

$$D_{(5,7)2} = \frac{1}{2} 0,49 + \frac{1}{2} 0,5$$

$$D_{(5,7)2} = \frac{0,99}{2} = 0,49$$

$$D_{(5,7)3} = \frac{n_5}{n_5 + n_7} d_{(5,3)} + \frac{n_7}{n_5 + n_7} d_{(7,3)}$$

$$D_{(5,7)3} = \frac{1}{2} 0,85 + \frac{1}{2} 0,84$$

$$D_{(5,7)3} = \frac{1,68}{2} = 0,84$$

$$D_{(5,7)4} = \frac{n_5}{n_5 + n_7} d_{(5,4)} + \frac{n_7}{n_5 + n_7} d_{(7,4)}$$

$$D_{(5,7)4} = \frac{1}{2} 1,38 + \frac{1}{2} 1,46$$

$$D_{(5,7)4} = \frac{2,83}{2} = 1,41$$

$$D_{(5,7)6} = \frac{n_5}{n_5 + n_7} d_{(5,6)} + \frac{n_7}{n_5 + n_7} d_{(7,6)}$$

$$D_{(5,7)6} = \frac{1}{2} 0,9 + \frac{1}{2} 0,94$$

$$D_{(5,7)6} = \frac{1,85}{2} = 0,92$$

$$D_{(5,7)8} = \frac{n_5}{n_5 + n_7} d_{(5,8)} + \frac{n_7}{n_5 + n_7} d_{(7,8)}$$

$$D_{(5,7)8} = \frac{1}{2} 1,05 + \frac{1}{2} 1,04$$

$$D_{(5,7)8} = \frac{2,09}{2} = 1,05$$

Dari penyelesaian di atas, maka hasil cluster tersebut dapat di sajikan dalam bentuk Tabel 4.37 berikut ini:

Tabel 4.10 Proximity Matriks Average Linkage

D_{Euclid}	(5,7)	1	2	3	4	6	8
(5,7)	0	1,34	0,5	0,84	1,45	0,94	1,05
1		0	1	0,68	1,55	1,81	0,93
2			0	0,49	1,3	1,24	0,76
3				0	1,58	1,39	1,08
4					0	1,62	1,19

6						0	1,55
8							0

Dan seterusnya dilakukan perhitungan jarak dari data-data yang telah di klusterisasikan dengan menggunakan metode Average Linkage. Proses pengklasteran dilakukan hingga tahap 5, kemudian akan didapatkan hasil akhir pengklasteran yang disajikan dalam Tabel berikut ini:

Tabel 4.9 Proximity Matriks Average Linkage final step cluster

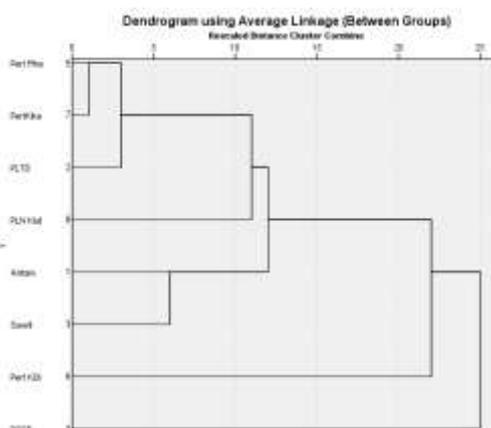
D_{Euclid}	(1,2,3,5,7,8)	4	6
(1,2,3,5,7,8)	0	1,59	1,81
4		0	1,62
6			0

Dapat dilihat bahwa jarak antara objek (1,2,3,5,7,8) dengan objek 4 memiliki jarak terdekat, yakni 1,587. Maka objek-objek tersebut di kelompokkan menjadi 1 cluster. Setelah objek-objek tersebut sudah dikelompokkan, maka sekarang tersisa 2 cluster. Sehingga terdapat 2 cluster yang terbentuk pada tahap terakhir yakni tahap 5, dimana cluster-cluster tersebut terdiri dari:

Cluster 1: PT DSSP Power Kendari (4)

Cluster 2: Industri Antam, PLTD Wangi-wangi, PT Merbaujaya Indahraya, PT DSSP Power Kendari, Pertamina Raha, Pertamina Kolaka, PLN Kolaka Utara (1,2,3,4,5,7,8).

Berikut ini merupakan hasil dendrogram dengan metode Complete Linkage dengan menggunakan Software IBM SPSS 26.



Gambar 4.7 Dendrogram Metode Average Linkage

4.7 Metode Non Hierarki

4.7.1 Penerapan Algoritma K-Means

Dalam melakukan klusterisasi dengan menggunakan metode K-Means, maka terlebih dahulu menentukan jumlah cluster (k) dan pusat (centroid). Untuk k peneliti menentukan nilai k = 2. Dan untuk menentukan pusat (centroid) awal, ditentukan dengan mengacak (random) dari objek yang sudah ada. Dalam

kasus ini, peneliti membuat 2 *centroid*. Untuk *centroid* 1, dan *centroid* 2, peneliti mengambil masing-masing dari objek ke-2 (PLTD Wangi-wangi) dan objek ke-8 (PLN Kolaka Utara).

1. Iterasi ke-1

Centroid 1(0,34 ; 0,03 ; 0,26 ; 0,32 ; 0,14)

Centroid 2(0,34 ; 0,02 ; 1 ; 0,11 ; 0,04)

Setelah menentukan jumlah *cluster* dan *centroid*, maka selanjutnya yaitu menghitung jarak masing-masing objek ke masing-masing *centroid* dengan menggunakan persamaan 2.14. sehingga diperoleh:

Untuk objek 1 (0,17 ; 0 ; 0,77 ; 1 ; 0)

Centroid M_1 (0,34 ; 0,03 ; 0,26 ; 0,31 ; 0,14)

Centroid M_2 (0,34 ; 0,02 ; 1 ; 0,11 ; 0,04)

$$DM_1 = \sqrt{(0,17 - 0,34)^2 + (0 - 0,03)^2 + (0,77 - 0,26)^2 + (1 - 0,31)^2 + (0 - 0,04)^2}$$

$$DM_1 = \sqrt{0,03 + 0 + 0,25 + 0,46 + 0}$$

$$= \sqrt{0,74}$$

$$= 0,86$$

$$DM_2 = \sqrt{(0,17 - 0,34)^2 + (0 - 0,02)^2 + (0,77 - 1)^2 + (1 - 0,11)^2 + (0 - 0,04)^2}$$

$$DM_2 = \sqrt{0,028 + 0 + 0,05 + 0,78 + 0}$$

$$= \sqrt{0,86}$$

$$= 0,93$$

Untuk objek 2 (0,34 ; 0,03 ; 0,26 ; 0,32 ; 0,15)

$$DM_1 = \sqrt{(0,34 - 0,34)^2 + (0,03 - 0,03)^2 + (0,26 - 0,26)^2 + (0,32 - 0,32)^2 + (0,15 - 0,15)^2} = 0$$

$$DM_2$$

$$= \sqrt{(0,34 - 0,34)^2 + (0,03 - 0,02)^2 + (0,26 - 1)^2 + (0,31 - 0,11)^2 + (0,15 - 0,04)^2}$$

$$DM_2 = \sqrt{0 + 0 + 0,53 + 0,04 + 0,01}$$

$$= \sqrt{0,58}$$

$$= 0,76$$

Dan seterusnya dilakukan perhitungan jarak objek ke-3 sampai objek ke-8. Kemudian akan didapatkan hasil perhitungan jarak setiap objek terhadap pusat *cluster* yang disajikan dalam Tabel berikut ini:

Tabel 4.11 Jarak Objek ke Masing-masing *Centroid*

Industri	C1	C2	Terdekat
Antam	0,87	0,93	C1
PLTD	0	0,76	C1
Sawit	0,49	1,08	C1
DSSP Power	1,3	1,19	C2
Pert Raha	0,49	1,04	C1
pert kendari	1,24	1,55	C1
Pert Kolaka	0,49	1,04	C1
PLN Kolut	0,76	0	C2

DDari Tabel 4.39 di atas, didapatkan keanggotaan sebagai berikut:

Cluster 1: Antam, PLTD Wangi-wangi, PT

Cluster 2: Merbaujaya Indahraya (Sawit),

Pertamina Raha, Pertamina

Kendari, dan Pertamina Kolaka.

PT DSSP Powwer Kendari dan

PLN Kolaka Utara.

2. Iterasi ke-2

Dari hasil iterasi pertama, maka sebelum melakukan iterasi ke-2, terlebih dahulu di lakukan pembaruan *Centroid*. *Centroid* baru diperoleh dengan menghitung nilai rata-rata dari tiap *cluster* yang bisa dilihat pada Tabel berikut ini :

Tabel 4.12 Pembaharuan *Centroid* untuk Iterasi ke-2

Cluster 1	Antam	0,17	0	0,771605	1	0
	PLTD	0,34	0,03	0,26	0,31	0,14
	Sawit	0,39	0,04	0,15	0,79	0,06
	Pertamina Raha	0,48	0,25	0	0	0,15
	Pertamina Kndri	1	1	0	0	0,17
	Pertamina Klaka	0,58	0,14	0	0	0,14
Rata-rata		0,49	0,24	0,19	0,35	0,11
Cluster 2	DSSP	0	0,63	0,9	0,02	1
	PLN Kolut	0,34	0,02	1	0,11	0,04
Rata-rata		0,17	0,32	0,95	0,06	0,52

Berdasarkan Tabel 4.40 di atas, maka dapat diambil *centroid* baru untuk iterasi ke-2 ini, yaitu:

Centroid 1 (0,49 ; 0,24 ; 0,19 ; 0,35 ; 0,11)

Centroid 2 (0,17 ; 0,32 ; 0,95 ; 0,06 ; 0,52)

Dan selanjutnya kita hitung dengan cara yang sama seperti pada iterasi pertama, bedanya adalah di nilai *centroid*. Dengan langkah-langkah seperti pada iterasi pertama sebelumnya, maka hasil iterasi ke-2 dengan *centroid* baru dapat dilihat pada Tabel berikut ini:

Tabel 4.13 Jarak Objek Terhadap *Centroid* pada Iterasi ke-2

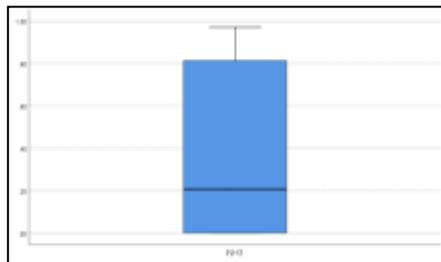
Dari Tabel 4.41 di atas, didapatkan keanggotaan dari iterasi ke-2 ini, yaitu sebagai berikut:

- Cluster 1:* Antam, PLTD Wangi-wangi PT
Cluster 2: Merbaujaya Indahraya (Industri Sawit), Pertamina Raha, Pertamina Kendari, dan Pertamina Kolaka.
 PT DSSP Power Kendari dan PLN Kolaka Utara.

Dari perhitungan iterasi ke-1 dan iterasi ke-2, diketahui bahwa setelah dicek, ternyata tidak terjadi pemindahan objek antar *cluster*. Dengan kata lain, *clustering* antara perhitungan iterasi ke-1 dan iterasi ke-2 mempunyai pengelompokan objek yang sama di setiap *cluster*. Maka pada metode *K-Means* ini, proses pengelompokannya berhenti pada iterasi ke-2. Sehingga untuk metode *K-Means* ini, terdiri dari dua *cluster*, yaitu:

- Cluster 1:* Antam dan PT Merbaujaya Indahraya (Industri Sawit)
Cluster 2: PLTD Wangi-wangi, PT DSSP Power Kendari, Pertamina Raha, Pertamina Kendari, Pertamina Kolaka, PLN Kolaka Utara.

Gambar 4.2. Box Plot Data 2



5. Penutup

5.1 Kesimpulan

Dari hasil analisis dan pembahasan, dapat diambil kesimpulan sebagai berikut:

1. Hasil analisis *cluster* dari 8 jenis industri perusahaan berdasarkan jenis polutan TSP, CO, NH₃, NO₂, dan SO₂ :
 - a. Untuk metode hierarki dapat di kelompokkan menjadi dua kelompok, yaitu *cluster* pencemar rendah dan *cluster* pencemar tinggi.
 - b. Untuk metode non hierarki juga dapat dikelompokkan menjadi dua *cluster*, yaitu *cluster* pencemar tinggi dan *cluster* pencemar rendah.
2. Ciri dari setiap *cluster* yang terbentuk yaitu:
 - a. Untuk metode hierarki:

Industri	C1	C2	Terdekat
Antam	0,95	1,13	C1
PLTD	0,27	0,88	C1
Sawit	0,49	1,22	C1
DSSP Power	1,33	0,59	C2
Pert Raha	0,4	1,07	C1
Pert Kendri	0,99	1,47	C1
Pert Klaka	0,42	1,11	C1
PLN Kolut	0,88	0,59	C2

1. *Cluster* pencemar rendah untuk metode *Complete Linkage* memiliki rata-rata nilai variabel (polutan) yang paling kecil, dan anggotanya adalah industri Antam, PLTD wangi-wangi, PT Merbaujaya Indahraya, PT DSSP Power Kendari, Pertamina Raha, Pertamina Kolaka, dan PLN Kolaka Utara. Sedangkan untuk metode *Average Linkage* anggotanya yaitu industri Antam, PLTD wangi-wangi, PT Merbaujaya Indahraya, Pertamina Raha, Pertamina Kendari, Pertamina Kolaka, dan PLN Kolaka Utara
2. *Cluster* pencemar tinggi untuk metode *Complete Linkage* memiliki rata-rata nilai variabel terbesar yang anggotanya adalah Pertamina Kendari. Sedangkan untuk metode *Average Linkage*, anggotanya yaitu PT DSSP Power Kendari.
 - a. Untuk metode non hierarki:
 1. *Cluster* pencemar tinggi, kedua polutannya memiliki rata-rata lebih besar dari *cluster* pertama. Anggotanya adalah PT DSSP Power Kendari dan PLN Kolaka Utara.
 2. *Cluster* pencemar rendah, keenam polutannya memiliki rata-rata lebih kecil dari *cluster* kedua. Anggotanya adalah industri Antam, PLTD Wangi-wangi, PT Merbaujaya Indahraya, Pertamina Raha, Pertamina Kendari, Pertamina Kolaka.
3. Untuk jumlah *cluster* yang dibentuk antara metode hierarki dan non hierarki ini yakni sama-sama terdiri dari 2 *cluster*. Perbedaannya adalah objek dalam *cluster* yang dikelompokkan. Dimana untuk *cluster* pertama pada metode hierarki terdiri dari PT DSSP Power Kendari dan Pertamina Kendari, sedangkan untuk *cluster* pertama pada metode non hierarki terdiri dari industri Antam dan PT Merbaujaya Indahraya.

5.2 Saran

Berdasarkan kesimpulan yang telah dikemukakan, maka peneliti mencoba untuk memberikan saran-saran sebagai berikut:

1. Untuk meningkatkan pengaruh polutan terhadap pencemaran udara pada sektor industri, maka harus diarahkan lagi ke polutan yang memberikan pengaruh besar ke sektor industri itu.
2. Untuk peneliti selanjutnya, agar menambah variabel lain yang mempengaruhi tingkat pencemaran udara pada sektor industri di Sulawesi Tenggara.
3. Peranan industri dapat ditingkatkan untuk relokasi industri pencemar udara ke kawasan atau zona industri.

Ucapan Terima kasih: Saya ucapkan terima kasih kepada pembimbing dan penguji saya yang telah memberikan saran dan dukungan dalam penyusunan tugas akhir ini.

Daftar pustaka

- [1] P. N. W. Imtitsal & J. Nugraha, (2017). Analisis Cluster Tingkat Kualitas Udara Ambien Jalan Raya di DIY 2015. *Prosiding SI MaNIs*, 1(1), 178–187.
- [2] U. Jannah, (2010). Perbandingan jarak Euclid dengan Jarak Mahalanobis pada Analisis Cluster Hirarki.
- [3] A. Johnson, X. Wang, F. Kong & M. Xue, (2011). Hierarchical cluster analysis of a convection-allowing ensemble during the hazardous weather testbed 2009 spring experiment. Part I: Development of the object-oriented cluster analysis method for precipitation fields. *Monthly Weather Review*, 139(12), 3673–3693.
- [4] J. Reynaldo, P. P. Adikara & R. C. Wihandika, (2020). Analisis Sentimen Mengenai Produk Toyota Avanza Menggunakan Metode Learning Vector Quantization Versi 3 (LVQ 3) dengan Seleksi Fitur Chi Square, Lexicon-Based Features serta Normalisasi Min-Max. *Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 4(3), 830–839.
- [5] S. F. Rodiyansyah, (2017). K-Means Dan Fuzzy C-Means Pada Analisis Data Polusi Udara Di Kota X. *Seminar Nasional Teknologi Informasi Dan Multimedia 2017*, 25–30.
- [6] S. Setiawan, (2019). Analisis Cluster Menggunakan Algoritma K-Means Untuk Mengetahui Kemampuan Pegawai di Bidang IT pada CV. Roxed LTD. 7, 341-347.
- [7] R. Sitepu & B. Gultom, (2011). Analisis Cluster Terhadap Tingkat Pencemaran Udara pada Sektor Industri di Sumatera Selatan. *Jurnal Penelitian Sains*, 14(3), 11-17.
- [8] C. Suhaeni & A. R. Kurnia, (2018). Perbandingan Hasil Pengelompokan menggunakan Analisis Cluster Berhirarki, K-Means Cluster, dan Cluster Ensemble (Studi Kasus Data Indikator Pelayanan Kesehatan Ibu Hamil). 14(1).
- [9] E. Widodo, N. N. Sari, I. Hidayati & F. Yubinas, (2018). Analisis Cluster Penderita Disabilitas Mental di Provinsi Daerah Istimewa Yogyakarta Tahun 2016 *KNPMP III 2018 ISSN : 2502-6526*. 577–586.
- [10] S. Wijaya, (2018). Pengelompokan kabupaten/Kota di Jawa Tengah Menurut Kualitas Udara Ambien Menggunakan Analisis K-Harmonic Mean Cluster. *Prosiding Seminar Nasional Statistika*, 1(1), 1689–1699.
- [11] N. Yanti & M. Ulfah, (2015). Aplikasi Jaringan Syaraf Tiruan Untuk Clustering Polutan Kimia Penyebab Pencemaran Udara. *JTT (Jurnal Teknologi Terpadu)*, 3(2), 68–74.