

PENGELOMPOKAN KUALITAS AIR DI KOTA PONTIANAK MENGGUNAKAN METODE WARD DAN SINGLE LINKAGE

Ayu Soraya

INTISARI

Analisis kluster merupakan salah satu analisis multivariat yang bertujuan untuk mengelompokan atau mengklasifikasi objek-objek menjadi beberapa kelompok berdasarkan kemiripan variabel-variabel yang diamati, sehingga diperoleh kemiripan objek dalam kelompok yang sama dibandingkan antar objek dari kelompok yang berbeda. Dalam penelitian ini metode Ward dan Single Linkage digunakan untuk mengelompokan kualitas air di Kota Pontianak. Berdasarkan perhitungan R-Squared diperoleh pengelompokan kualitas air menggunakan Metode Single Linkage adalah metode terbaik dengan nilai R-Squared 0,970 dengan kluster yang terbaik dikluster ketiga yang beranggotakan satu lokasi sampel air yaitu jalan Selat Sumba II.

Kata Kunci: Analisis kluster, metode ward, metode single linkage.

PENDAHULUAN

Analisis kluster merupakan salah satu analisis multivariat yang bertujuan untuk mengelompokan atau mengklasifikasi objek-objek menjadi beberapa kelompok berdasarkan kemiripan variabel-variabel yang diamati, sehingga diperoleh kemiripan objek dalam kelompok yang sama dibandingkan antar objek dari kelompok yang berbeda [1]. Metode pengelompokan yang bisa digunakan dalam analisis kluster hirarki yaitu metode pautan tunggal (*Single Linkage*), metode pautan lengkap (*Complete Linkage*), metode antar pusat (*Centroid Linkage*), metode pautan rata-rata (*Average Linkage*) dan metode Ward (*Ward's Method*). Metode Ward merupakan suatu metode pembentukan kluster yang didasari oleh hilangnya informasi akibat penggabungan objek menjadi kluster. Hal ini diukur dengan menggunakan jumlah total dari deviasi kuadrat pada *mean cluster* untuk setiap pengamatan. Analisis kluster hirarki pada metode *Single Linkage* proses pengelompokan menggunakan jarak minimum antar kelompok. Proses pengelompokan diawali dengan menemukan dua objek yang mempunyai jarak minimum dan untuk selanjutnya objek tersebut menjadi satu kelompok.

Air merupakan kebutuhan dasar yang sangat penting bagi makhluk hidup. Pada kehidupan sehari-hari, air digunakan untuk memenuhi kebutuhan hidup manusia serta makhluk hidup lainnya. Kualitas air sangat mempengaruhi tingkat kesehatan makhluk hidup. Perubahan kondisi kualitas air berdampak dari penggunaan lahan yang ada. Jumlah penduduk yang semakin meningkat di kota Pontianak berpengaruh terhadap penggunaan lahan. Perlindungan terhadap kualitas air sangat penting dilakukan untuk memenuhi baku mutu yang telah ditetapkan.

Penelitian ini bertujuan untuk menganalisis pengelompokan kualitas air di Kota Pontianak berdasarkan karakteristik titik lokasi sampel air yaitu kekeruhan (x_1), warna (x_2), TDS (x_3), PH (x_4), besi (x_5), fluorida (x_6), kesadahan (x_7), nitrat (x_8), nitrit (x_9), deterjen (x_{10}), DO (x_{11}), BOD₅ (x_{12}) dan COD (x_{13}) dan membandingkan hasil antara metode Ward dan *Single Linkage*.

Langkah-langkah dalam penelitian ini adalah pengumpulan data, input data, data yang digunakan adalah sampel air permukaan yang diambil dari 42 titik lokasi di Kota Pontianak. Sebelum dilakukan analisis kluster dengan menggunakan metode Ward dan *Single Linkage* terlebih dahulu dilakukan standarisasi data. Kemudian dilakukan uji kecukupan data secara keseluruhan yang digunakan adalah uji *Kaiser-Mayer-Olkin* (KMO). Setelah itu, melakukan uji multikolinearitas dengan melihat nilai

varians inflation factor (VIF), jika nilai $VIF < 10$ maka tidak terindikasi multikolinearitas dan dapat dilanjutkan dengan langkah selanjutnya. Sebaliknya jika nilai $VIF > 10$ maka terindikasi multikolinearitas dan perlu dilakukan transformasi data dengan cara mengeluarkan variabel independen yang berkorelasi. Menentukan ukuran kemiripan dan ketidakmiripan antar dua objek menggunakan pengukuran jarak kuadrat *Euclidean*. Pengukuran jarak kuadrat *Euclidean* dilakukan untuk melihat seberapa mirip atau berbeda objek tersebut. Setelah melakukan pengukuran jarak, langkah selanjutnya yang dilakukan analisis kluster menggunakan metode *Ward* dan *Single Linkage* dan membandingkan metode tersebut dengan melihat hasil perhitungan nilai *R-Squared* (R^2).

ANALISIS KLASTER

Analisis kluster merupakan teknik multivariat yang digunakan untuk mengelompokkan objek menjadi kelompok-kelompok yang lebih kecil dimana setiap kelompok berisi objek yang mirip satu sama lain.

1. Standarisasi Data

Proses standarisasi data dilakukan apabila diantara variabel-variabel yang diteliti terdapat variabilitas satuan. Perbedaan satuan yang mencolok dapat mengakibatkan perhitungan pada analisis kluster menjadi tidak valid. Proses standarisasi dilakukan dengan transformasi pada data asli sebelum dianalisis lebih lanjut. Transformasi dilakukan terhadap variabel yang relevan kedalam bentuk z_{score} sebagai berikut:

$$z = \frac{x - \bar{x}}{s} \quad (1)$$

dengan

x = nilai data

s = standar deviasi

\bar{x} = nilai rata-rata

2. Mengukur Kemiripan Antar Objek

Sesuai dengan prinsip kluster yaitu mengelompokkan objek yang memiliki kemiripan, maka dilakukan pengukuran seberapa jauh ada kesamaan kesamaan antar objek [2]. Terdapat beberapa macam ukuran jarak yang biasa dipakai dalam analisis kluster, diantaranya yaitu jarak *Euclidean*. Jarak *Euclidean* adalah jarak antara dua objek dari p dimensi pengamatan. Terdapat objek pertama yang diamati adalah $i = (x_{i1}, x_{i2}, \dots, x_{ip})$ dan $j = (x_{j1}, x_{j2}, \dots, x_{jp})$ jarak kuadrat *Euclidean* dapat di hitung dengan rumus:

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

dengan

d_{ij} = jarak kuadrat *Euclidean* antar objek ke- i dengan objek ke- j

p = jumlah variabel

x_{ik} = nilai atau data dari objek ke- i pada variabel ke- k

x = nilai atau data dari objek ke- j pada variabel ke- k

METODE WARD DAN SINGLE LINKAGE

Metode *Ward* berusaha untuk meminimalkan variasi antar objek yang ada dalam satu kluster. Hal ini diukur dengan menggunakan jumlah total dari deviasi kuadrat pada *mean cluster* untuk setiap pengamatan. *Error sum of squares* (*SSE*) digunakan sebagai fungsi objektif [4]. *SSE* hanya dapat dihitung jika kluster memiliki elemen lebih dari satu objek.

$$SSE = \sum_{i=1}^n (x_i - \bar{x})(x_j - \bar{x}) \quad (3)$$

dengan

x_i = vektor kolom yang berisikan nilai objek i

x_j = vektor kolom yang berisikan nilai objek j

\bar{x} = vektor kolom yang entrinya rata-rata nilai objek dalam kluster

n = banyak objek

Metode *single linkage* dimulai dengan mengelompokan dua objek yang dipisahkan dengan jarak paling dekat pada kluster pertama dan seterusnya. Metode ini dikenal pula dengan nama pendekatan tetangga terdekat [5]. Secara formal dua buah jarak antara objek u dan objek v misalnya $d(u,v)$ didefinisikan sebagai berikut:

$$d(u,v) = \min\{d(x_i,x_j), x_i \in u \text{ dan } x_j \in v\} \tag{4}$$

dengan

d_{uv} = jarak minimum antara kluster u dan v

STUDI KASUS

Sebelum proses perhitungan data dengan metode kluster, terlebih dahulu dilakukan pengumpulan data. Pengumpulan data dilakukan secara tidak langsung. Data yang digunakan adalah sampel air permukaan yang diambil dari 42 titik lokasi di kota Pontianak. Variabel yang diamati dalam penelitian ini yaitu kekeruhan (x_1), warna (x_2), TDS (x_3), PH (x_4), besi (x_5), fluorida (x_6), kesadahan (x_7), nitrat (x_8), nitrit (x_9), deterjen (x_{10}), DO (x_{11}), BOD₅ (x_{12}) dan COD (x_{13}). Selanjutnya dicari hasil analisis statistik deskriptif melalui bantuan *software SPSS 20*. Hasil analisis statistik deskriptif tersebut dapat dilihat pada Tabel 1 berikut :

Tabel 1 Statistik deskriptif

Variabel	Minimum	Maksimum	Rata-rata	Standar deviasi
x_1 (NTU)	2,21	92,60	21,44	16,37
x_2 (Pt.Co)	22,00	990,00	365,38	329,55
x_3 (mg/l)	17,20	156,30	74,14	37,85
x_4 (Ph)	3,88	7,94	6,89	1,05
x_5 (mg/l)	0,08	6,28	1,91	1,35
x_6 (mg/l)	0,18	1,53	0,42	0,21
x_7 (mg/l)	8,00	68,00	38,83	13,36
x_8 (mg/l)	0,10	9,00	1,24	1,42
x_9 (mg/l)	0,00	0,56	0,22	0,25
x_{10} (mg/l)	0,00	0,01	0,0019	0,00153
x_{11} (mg/l)	1,01	5,25	2,65	1,23
x_{12} (mg/l)	7,62	50,50	18,57	8,65
x_{13} (mg/l)	31,05	126,86	80,99	30,91

Selanjutnya dicari nilai KMO, berdasarkan output SPSS diperoleh nilai KMO yang disajikan pada Tabel 2.

Tabel 4.2 KMO

Keterangan	Nilai
Kaiser-Meyer-Olkin Measure of Sampling Adequacy	0,619

Kaiser Meyer Olkin (KMO) adalah indeks perbandingan jarak antara koefisien korelasi dengan koefisien korelasi parsialnya. Jika jumlah kuadrat koefisien korelasi parsial di antara seluruh pasangan variabel bernilai kecil dibandingkan dengan jumlah kuadrat koefisien korelasi, maka akan menghasilkan nilai *KMO* mendekati 1. Nilai *KMO* dianggap mencukupi jika lebih dari 0,5. Hasil penelitian menunjukkan bahwa nilai *KMO* sebesar 0,619. Dengan demikian persyaratan *KMO* > 0,5 terpenuhi. Sehingga jumlah data cukup untuk difaktorkan.

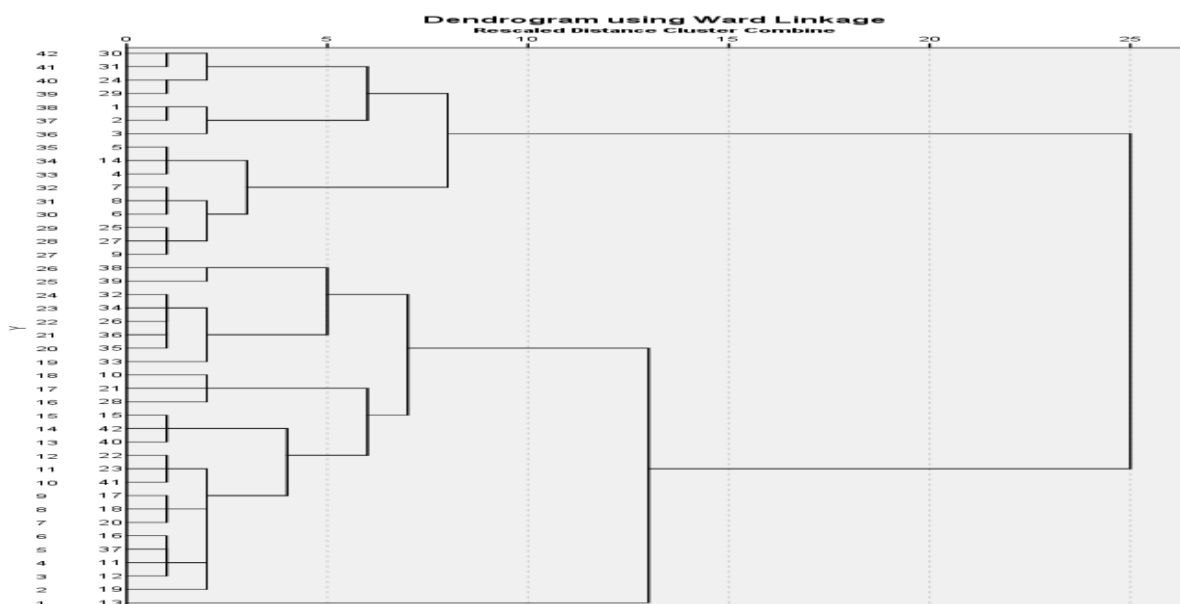
Setelah itu, melakukan uji multikolinieritas. Berdasarkan output SPSS 20 diperoleh nilai *VIF* yang disajikan pada Tabel 3.

Tabel 3 Multikolinieritas

Variabel	VIF
x ₁ (NTU)	1,784
x ₂ (Pt.Co)	8,985
x ₃ (mg/l)	8,396
x ₄ (Ph)	4,668
x ₅ (mg/l)	2,267
x ₆ (mg/l)	3,636
X ₇ (mg/l)	3,930
x ₈ (mg/l)	3,946
x ₉ (mg/l)	1,997
x ₁₀ (mg/l)	1,899
x ₁₁ (mg/l)	1,821
x ₁₂ (mg/l)	1,744
x ₁₃ (mg/l)	5,559

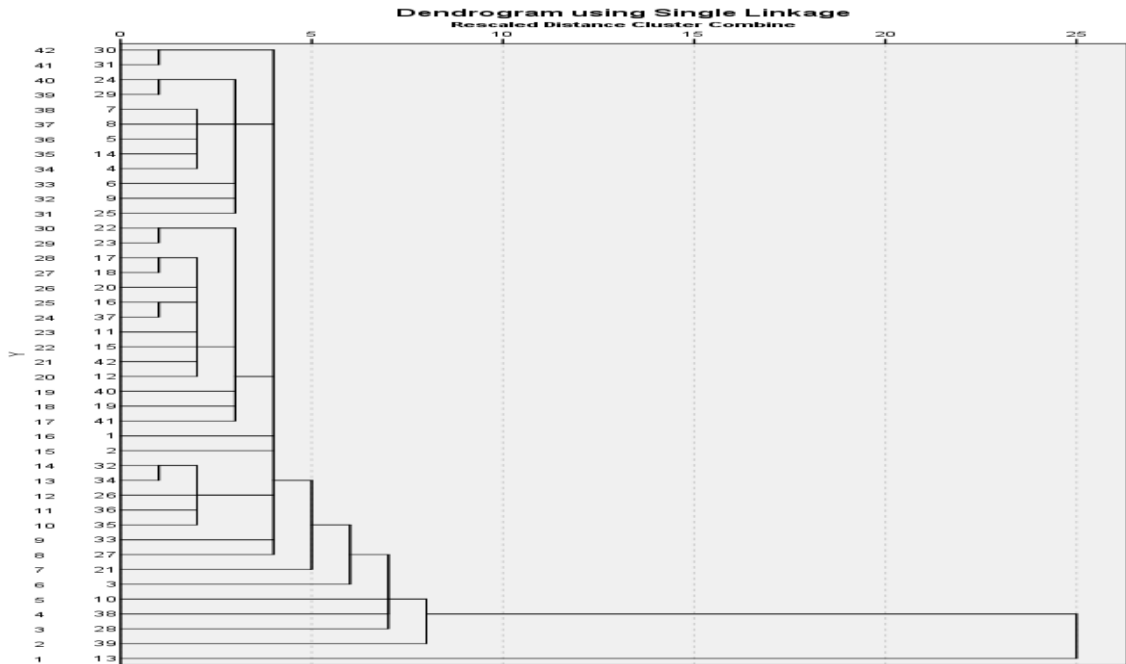
Multikolinieritas terindikasi apabila nilai *VIF* > 10. Salah satu cara yang dapat dilakukan bila terjadi multikolinieritas adalah dengan mengeluarkan variabel yang berkorelasi dalam model. Berdasarkan Tabel 3, diketahui bahwa nilai *VIF* pada tiap variabel bernilai kurang dari 10. Maka dapat diketahui bahwa tidak terjadi multikolinieritas antar variabel-variabel tersebut.

Proses pengelompokan kualitas air pada 42 titik lokasi sampel di kota Pontianak dengan metode *Ward* dan *Single Linkage* menggunakan bantuan SPSS. Hasil pengelompokan kualitas air berdasarkan Variabel yang diamati dalam penelitian ini dapat dilihat dalam bentuk dendrogram. Hasil dendrogram dengan menggunakan metode *Ward* dapat dilihat di Gambar 1.



Gambar 1 dendrogram metode *Ward*

Hasil dendrogram dengan menggunakan metode *Single Linkage* dapat dilihat di Gambar 2.



Gambar 2 dendrogram metode *Single Linkage*

Selanjutnya menentukan metode terbaik dengan membandingkan nilai *R-Squared*. Hasil dari perhitungan nilai *R-Squared* menggunakan metode *Ward* dan metode *Single Linkage* disajikan pada Tabel 4.

Tabel 4 Nilai *R-Squared* pada *Ward* dan *Single Linkage*

Metode	Klaster	SS_b	SS_t	R^2	Rata-rata R^2
<i>Ward</i>	1	85,182	139,441	0,611	0,535
	2	35,494	75,470	0,470	
	3	51,582	250,115	0,206	
	4	58,106	68,050	0,854	
<i>Single Linkage</i>	1	2,523	427,523	0,006	0,694
	2	58,106	68,050	0,854	
	3	19,481	20,098	0,970	
	4	16,469	17,406	0,946	

Keterangan : SS_b adalah jumlah kuadrat antar kelompok dan SS_t adalah jumlah kuadrat total.

Untuk menggambarkan isi klaster pada Tabel 4, diperoleh interpretasi klaster sebagai berikut :

- 1) Pada pengelompokan kualitas air menggunakan Metode *Ward*, dipilih satu klaster yang terbaik diklaster keempat dengan nilai R^2 0,854 yang beranggotakan satu lokasi sampel air nomor 13 pada Gambar 1 yaitu jalan Dr Wahidin Pertigaan Ujung Pandang Depan Masjid.
- 2) Pada pengelompokan kualitas air menggunakan Metode *Single Linkage*, dipilih satu klaster yang terbaik diklaster ketiga dengan nilai R^2 0,970 yang beranggotakan satu lokasi sampel air nomor 28 pada Gambar 2 yaitu jalan Selat Sumba II.

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang dilakukan terhadap data kualitas air di Kota Pontianak, diperoleh kesimpulan sebagai berikut :

1. Pada pengelompokan kualitas air menggunakan Metode *Ward*, dipilih satu klaster yang terbaik diklaster keempat dengan nilai R^2 0,854 yang beranggotakan satu lokasi sampel air nomor 13 pada Gambar 1 yaitu jalan Dr Wahidin Pertigaan Ujung Pandang Depan Masjid.
2. Pada pengelompokan kualitas air menggunakan Metode *Single Linkage*, dipilih satu klaster yang terbaik diklaster ketiga dengan nilai R^2 0,970 yang beranggotakan satu lokasi sampel air nomor 28 pada Gambar 2 yaitu jalan Selat Sumba II.
3. Berdasarkan nilai R -Squared pada metode klaster, dapat disimpulkan bahwa metode *Single Linkage* menghasilkan klaster lebih baik dibandingkan metode *Ward* karena nilai R metode² *Single Linkage* lebih besar.

DAFTAR PUSTAKA

- [1] Supranto, J., 2010, *Analisis Multivariat Arti dan Interpretasi*, Rineka Cipta, Jakarta.
- [2] Simamora, B., 2005, *Analisis Multivariat Pemasaran Edisi Pertama*, Jakarta:PT.Gramedia Pustaka Tama.
- [3] Gunawan, Iman., 2016, *Penerapan Statistika Inferensial*, Rajawali Pers, Jakarta.
- [4] Hair, J.F., Black, W.C., Babin, B.J., and Anderson, R.E, 2010. *Multivariate Data Analysis*, New Jersey: Upper Saddle River.
- [5] Gudono., 2011, *Analisis Data Multivariat*, Edisi Pertama, Yogyakarta: BPFE.

AYU SORAYA : Jurusan Matematika FMIPA Untan, Pontianak

ayusoraya@student.untan.ac.id
