

SISTEM IDENTIFIKASI PERSEBARAN PECEMARAN AIR OLEH LIMBAH DI INDONESIA MENGGUNAKAN AVERAGE LINKAGE DAN K-MEAN CLUSTER

Lalu Mutawalli¹, Mohammad Taufan Asri Zaen², Indi Febriana Suhriani³

¹Jurusan Teknik Informatika, Universitas Islam Indonesia, Yogyakarta

²Prodi Sistem Informasi, STMIK Lombok, Praya

³Prodi Matematika, Universitas Gadjah Mada, Yogyakarta

¹laluallistilo@gmail.com, ²opanzain@gmail.com, ³indifebriana94@gmail.com

Abstract

Water contamination is a problem that is always difficult to resolved. One of the main sources that causes water contamination is waste caused by human activities. The needed for a system that can analyze the data of water contamination sources. The main cause of water contamination that became variables in this study are family, factory and other waste. The method of Cluster and Average Linkage is used to analyze hierarchical data. The results of Cluster analysis hierarchically divided into three provincial groups based on the population distribution of waste. The first group is the Province of Nanggroe Aceh Darussalam, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, Bangka Belitung, DKI Jakarta, DI Yogyakarta, East Java, Banten, Bali, NTB, NTT, East Kalimantan, North Sulawesi, Central Sulawesi, South Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, North Maluku, West Papua and Papua. The second group consists of West Java and Central Java. In the third group occupied by West Kalimantan, Central Kalimantan and South Kalimantan. The source of water contamination, namely family waste, dominates the second group or it can be said that the province classified as the second group is dominated by family waste. The source of factory waste water contamination that dominates in the third group or it can be said that the provinces classified in the third group are dominated by factory waste as one of the most important sources of water contamination. The first group consisted of 28 members or 28 provinces, the second group had 2 members, while the third group consisted of 3 members. The first group has a source of water contamination, the most important of which is based on the indicators that are seen to be stable for the three indicators. The main source of water contamination based on the three indicators studied for the second group is dominated by family waste and other wastes. Whereas for the third group is dominated by factory waste.

Keywords: Water Contamination, Waste, Cluster Analysis, Average Linkage

Abstrak

Pencemaran air merupakan permasalahan yang selalu sulit terselesaikan, salah satu sumber penyebab utama pencemaran air adalah limbah yang diakibatkan oleh aktivitas manusia. Diperlukan suatu sistem yang dapat melakukan analisis dari data sumber pencemaran air. Penyebab utama pencemaran air yang menjadi variabel pada penelitian ini adalah limbah keluarga, pabrik, dan limbah lainnya. Metode cluster dan Average Linkage digunakan untuk menganalisa data secara hierarki. Hasil analisis cluster secara hierarki dibagi tiga kelompok Provinsi berdasarkan populasi persebaran limbah. Kelompok pertama yaitu Provinsi Nanggroe Aceh Darussalam, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, DKI Jakarta, DI Yogyakarta, Jawa Timur, Banten, Bali, NTB NTT, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat dan Papua. Kelompok kedua terdiri dari Jawa Barat dan Jawa Tengah. Pada kelompok ketiga ditempati oleh Kalimantan Barat, Kalimantan Tengah dan Kalimantan Selatan. Sumber pencemaran air yakni limbah keluarga mendominasi pada kelompok kedua atau dapat dikatakan bahwa provinsi yang tergolong dalam kelompok kedua didominasi oleh limbah keluarga. Sumber pencemaran air limbah pabrik yang mendominasi pada adalah kelompok ketiga atau dapat dikatakan bahwa provinsi yang tergolong dalam kelompok ketiga didominasi oleh limbah pabrik sebagai salah satu sumber pencemaran air yang paling utama. Kelompok pertama terdiri dari 28 anggota atau 28 provinsi, kelompok kedua memiliki anggota sebanyak 2 anggota, sedangkan kelompok ketiga terdiri dari 3 anggota. Kelompok pertama memiliki sumber pencemaran air yang paling utama berdasar indikator yang dilihat adalah cenderung stabil untuk ketiga indikator tersebut. Sumber pencemaran air yang paling utama berdasarkan dari ketiga indikator yang diteliti untuk kelompok kedua didominasi oleh limbah keluarga dan limbah lainnya. Sedangkan untuk kelompok ketiga didominasi oleh limbah pabrik.

Kata Kunci: Pencemaran Air, Limbah, Analisis Cluster, Average Linkage

1. Pendahuluan

Sebagai Negara kepulauan yang terbesar di dunia Indonesia dengan potensi perairan yang sangat luas terbentang, sehingga dalam nawa cita pemerintah ingin membuat poros maritim yang memiliki produktifitas yang tinggi sehingga hal ini akan memberikan dampak perekonomian yang baik bagi rakyat Indonesia. Secara geografis Indonesia memiliki wilayah yang sangat di daratan dan di lautan (Wahyono, 2009) laut Indonesia memiliki kekayaan yang sangat kaya, meliputi kekayaan hayati dan nabati. Hal ini menunjukkan bahwa sesungguhnya apabila perairan Indonesia mampu di kelola dengan baik maka akan memberikan kesejahteraan bagi rakyatnya.

Meskipun Indonesia memiliki potensi yang sangat besar di perairannya namun belum mampu memberikan kontribusi yang maksimal terhadap negara, dalam tulisannya (Panjaitan M., 2000) menuliskan Indonesia memiliki potensi lautan yang sangat besar akan tetapi tidak mampu di kelola dengan baik, sehingga populasi kemiskinan masih belum terbantu dalam pengurangan populasinya melalui sektor ini. Selain itu pencemaran air di Indonesia sangat besar, sebagian besar diakibatkan oleh aktifitas manusia yang meninggalkan limbah pemukiman, limbah pertanian, limbah industri dan limbah pertambangan. Hal ini memiliki efek yang sangat besar terhadap perairan Indonesia, karena dampaknya akan mempengaruhi pertumbuhan baik di daratan maupun di lautan.

Persebaran limbah di Indonesia hampir merata disetiap wilayah provinsi dan masih minim penanganannya, sehingga dalam penelitian ini akan dilakukan pengelompokan wilayah provinsi mana yang menyumbang limbah paling banyak di Indonesia, limbah keluarga secara memiliki skala yang paling besar hal ini akan menyebabkan ekspor nutrisi laut ke pesisir pantai yang lebih tinggi (Amin, Kroeze, & Strokal, 2017) meningkatkan resiko terhadap keberlangsungan hidup biota laut dan manusia sebagai konsumen hasil laut. Sehingga perlu disadari oleh setiap pemerintah daerah terhadap populasi limbah sehingga kebijakan pusat dan daerah akan memberikan penanganan yang intens terhadap limbah di Indonesia.

Dalam penelitian ini akan dilakukan pengelompokan terhadap objek data limbah, metode yang akan digunakan adalah k-mean cluster dimana telah banyak digunakan dalam berbagai kasus, pemanfaatan k-mean memberikan peningkatan dalam pemodelan distribusi spasial lokasi sampel dan viogram (Kapageridis, 2015) sehingga sebelum melakukan suatu pemodelan proses pengelompokan data akan memberikan suatu acuan yang lebih dinamis dalam pembuatan model. Penggunaan k-mean untuk mendeteksi area konsumsi listrik rumah tangga (Chevez, Barbero,

Martini, & Discoli, 2017) dalam penelitiannya mengidentifikasi faktor-faktor sosio demografis utama yang berdampak pada permintaan listrik dan lokasi geografis dari area konsumsi listrik. Sehingga k-mean memiliki kemampuan yang relevan jika digunakan untuk melakukan proses cluster terhadap kasus persebaran besaran limbah di Indonesia.

Pada analisis cluster model hirarki dikenal sebagai suatu metode dalam memodelkan kedekatan, kedekatan suatu kelompok dalam pemodelan hirarki, dalam hal ini kedekatan akan diukur berdasarkan rata-rata jarak suatu objek ini dikenal dengan metode *average linkage*, pada penelitian (Aris, Ayu, & Sari, 2016) *average linkage* metode ini mampu memberikan hasil yang baik dalam pengelompokan desa berdasarkan indikator pendidikannya. Sehingga menjadi salah satu acuan dalam melakukan pencarian jarak dalam kasus pengelompokan limbah yang mencemari perairan di Indonesia.

2. Metode Penelitian

Pada penelitian digunakan data sebaran limbah di Indonesia, data ini bersumber dari Badan Nasional penanggulangan Bencana (BNPB) dimana indikator yang digunakan adalah, provinsi dan sumber pencemaran air yang paling utama adalah limbah keluarga, limbah pabrik, dan limbah lainnya. Terdapat 33 wilayah provinsi dengan jumlah data persebarannya.

A. Langkah-langkah analisis data sebagai berikut:

1. Melakukan standarisasi data.
2. Melakukan pengecekan korelasi
3. Jika ada korelasi yang signifikan antara variabel dilakukan analisis komponen utama untuk membentuk komponen-komponen yang saling bebas. Selanjutnya skor komponen yang diperoleh dipandang sebagai variabel untuk analisis lebih lanjut. Jika tidak ada korelasi signifikan maka dilanjutkan langkah 4.
4. Mengukur kesamaan objek (similarity) menggunakan jarak *Euclidean* dengan rumus:

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

5. Menjelaskan setiap anggota kelompok, penjelasan kelompok menggunakan metode *average linkage* dengan rumus:

$$d(i, j)k = \text{average}(d_{ik}, d_{jk}) \quad (2)$$

6. Pengelompokan terhadap setiap objek proses pengelompokan data kedalam satu

cluster dapat dilakukan dengan menghitung jarak terdekat dari suatu data ke sebuah titik *centroid*. Rumus untuk menghitung jarak tersebut adalah:

$$d(x_i, x_j) = (|x_{i1} - x_{j1}|^g + |x_{i2} - x_{j2}|^g + \dots + |x_{ip} - x_{jp}|^g)^{1/g} \quad (3)$$

Pembaharuan suatu titik *centroid* dapat dilakukan dengan rumus berikut:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad (4)$$

B. Konsep Analisis Cluster

Analisa klaster memegang peran penting dalam pengklasifikasian objek. Bergantung pada aplikasinya, objek berupa sinyal, pelanggan, pasien, berita, tanaman dan lain-lain. Teknik klastering adalah teknik nonparametric yang sangat banyak diaplikasikan dalam kasus nyata (Santosa, 2007) teknik klaster dapat dikelompokkan kedalam dua kelompok besar yaitu, *partitioning cluster* dan *hierarchical cluster*.

1. Konsep Metode Hierarki

Dalam proses penggabungan kelompok dengan metode hierarki selalu diikuti dengan perbaikan matriks jarak. Suatu fungsi disebut jarak, jika mempunyai sifat tak negatif ($d_{ij} \geq 0$) dan ($d_{ij} = 0$), jika $i = j$ simetris ($d_{ij} = d_{ji}$), panjang salah satu sisi segitiga selalu lebih kecil atau sama dengan jumlah dua sisi yang lain ($d_{ij} \leq d_{ik} + d_{jk}$).

Beberapa macam jarak yang biasa dipakai di dalam analisis kelompok, yakni:

a. Jarak Euclidean

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (5)$$

b. Jarak Manhattan

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}| \quad (6)$$

c. Jarak Pearson

$$d_{ij} = \sqrt{\sum_{k=1}^p \frac{(x_{ik} - x_{jk})^2}{\text{var}(x_k)}} \quad (7)$$

d. Jarak Korelasi

$$d_{ij} = 1 - r_{ij} \quad (8)$$

e. Jarak Korelasi Mutlak

$$d_{ij} = 1 - |r_{ij}| \quad (9)$$

Dalam metode hirarki cluster terdapat dua tipe dasar yaitu, *agglomerative* (pemusatan) dan *divisive* (penyebaran). Dalam metode *agglomerative*, setiap obyek atau observasi dianggap sebagai sebuah cluster tersendiri. Dalam tahap selanjutnya dua cluster yang mempunyai kemiripan digabungkan menjadi sebuah cluster baru demikian seterusnya. Sebaliknya dalam metode *divisive* kita beranjak dari sebuah cluster besar yang terdiri dari semua obyek atau observasi. Selanjutnya, obyek atau observasi yang paling tinggi nilai ketidakmiripannya kita pisahkan demikian seterusnya. Dalam *agglomerative* ada lima metode yang cukup terkenal, yaitu : *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Ward's Method*, *Centroid Method*.

Single Linkage, prosedur ini didasarkan pada jarak terkecil. Jika dua obyek terpisah oleh jarak yang pendek maka kedua obyek tersebut akan digabung menjadi satu cluster dan demikian seterusnya. *Complete Linkage*, berlawanan dengan *Single Linkage* prosedur ini pengelompokkannya berdasarkan jarak terjauh. *Average Linkage*, prosedur ini hampir sama dengan *Single Linkage* maupun *Complete Linkage*, namun kriteria yang digunakan adalah rata-rata jarak seluruh individu dalam suatu cluster dengan jarak seluruh individu dalam cluster yang lain. *Ward's Method*, jarak antara dua cluster dalam metode ini berdasarkan total *sum of square* dua cluster pada masing-masing variabel. *Centroid Method*, jarak antara dua cluster dalam metode ini berdasarkan jarak *centroid* dua cluster yang bersangkutan.

3. Desain Dan Implementasi

Penelitian ini terbagi menjadi beberapa tahapan, dimana sebelum melakukan proses indentifikasi kedekatan dan pengelompokan terhadap objek hal yang pertama dilakukan adalah:

Deskripsi data

Provinsi di Indonesia sebanyak 33 dengan beberapa indikator yaitu sumber pencemaran, dimana pada sumber pencemaran terdiri dari limbah keluarga, limbah pabrik, dan limbah lainnya berkorelasi secara signifikan. Terlebih dahulu dilakukan analisis komponen utama untuk memperoleh skor komponen yang saling bebas. Skor komponen-skor komponen yang selanjutnya dianalisis dengan analisis untuk mengelompokkan provinsi di Indonesia berdasarkan persebaran limbah.

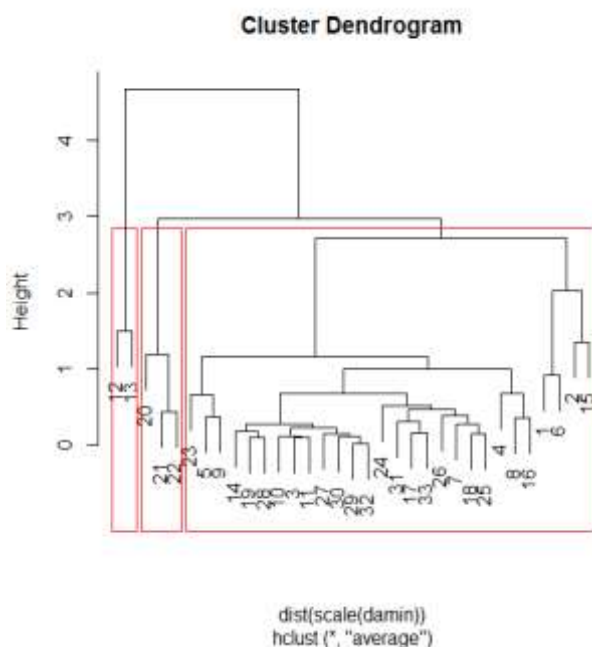
1. Analisis average linkage

```
Call:
hclust(d = dist(scale(damin)), method = "ave")

Cluster method : average
Distance       : euclidean
Number of objects: 33
```

Gambar 1 Jarak pada objek

Berdasarkan pada hasil pada gambar 1 terlihat bahwa metode analisis *clustering* secara hierarki yang digunakan dalam penelitian ini adalah metode "Average Linkage". Hal ini dapat dilihat dari penggunaan *script* untuk mendefinisikan metode tersebut, yakni "method=ave". Pada gambar 1 dapat dilihat pula bahwa jarak yang digunakan dalam penggunaan analisis *clustering* untuk studi kasus pada penulis kali ini adalah jarak *euclidean* dengan banyak observasi dari setiap variabel adalah 33. Angka 33 yang menunjukkan banyaknya data observasi yang digunakan merujuk pada provinsi di Indonesia pada tahun 2011.



Gambar 2 Dendrogram jarak antar objek

Berdasarkan pada gambar 2 dapat dilihat bahwa terdapat 28 observasi atau 28 Provinsi pada kelompok pertama yakni Provinsi yang terletak pada urutan 1-11, 14-19, dan 23-33. Kelompok kedua diisi dengan observasi pada urutan ke-12 dan ke-13. Sedangkan kelompok ketiga di isi dengan observasi ke-20 hingga observasi ke-22.

Tabel 1 perhitungan nilai rata-rata setiap variabel

Kelompok 1	63.03571	44.96429	50.60714286
Kelompok 2	264.5	418	215.5
Kelompok 3	81.66667	65.33333	314

Tabel 1 merupakan hasil perhitungan nilai rata-rata untuk setiap variabel dari masing-masing kelompok. Penghitungan nilai rata-rata tersebut maksudnya untuk mengetahui variabel yang mana saja yang mendominasi di setiap kelompok. Informasi tersebut dapat diperoleh dengan membandingkan nilai rata-rata suatu variabel dari ketiga kelompok yang telah dibentuk. Nilai rata-rata yang tertinggi dari suatu kelompok untuk masing-masing variabel, maka dikatakan bahwa kelompok ke- i , dimana $i = 1, 2, 3$ di dominasi oleh variabel ke- j , dimana $j = 1, 2, 3$. Berdasarkan Tabel 1 maka dapat dikatakan bahwa untuk kelompok 1 tidak didominasi oleh variabel, karena keseluruhan variabel independen apapun atau dapat dikatakan dikatakan bahwa sumber pencemaran air yang berasal dari tiga limbah untuk provinsi-provinsi yang berada di kelompok 1 masih lebih rendah dibandingkan dengan kelompok yang didominasi oleh limbah-limbah tersebut.

Berdasarkan Tabel 1 dapat diketahui bahwa untuk variabel limbah keluarga dan limbah lainnya mendominasi kelompok 2, hal ini berarti dapat dikatakan bahwa provinsi-provinsi yang berada di kelompok kedua memiliki tingkat pencemaran air yang berasal dari kedua jenis limbah tersebut lebih tinggi dibandingkan dengan provinsi yang terdapat pada kelompok lainnya. Sedangkan untuk variabel limbah pabrik mendominasi terhadap kelompok ketiga, yang artinya bahwa provinsi pada kelompok ketiga mempunyai pencemaran air yang berasal dari limbah pabrik yang lebih tinggi dibandingkan dengan provinsi di kelompok lainnya.

2. Analisis dengan metode k-means

	Cluster		
	1	2	3
Zscore (Limbah_Keluarga)	.14120	2.86033	-.88007
Zscore(Limbah_Pabrik)	-.01330	3.78654	-.55480
Zscore(Limbah_Lainnya)	3.19438	1.62056	-.80313

Gambar 3 initial cluster center

Berdasarkan pada gambar 3 yang menampilkan *Tabel Initial Cluster Centers* merupakan tampilan pertama proses *clustering* data sebelum dilakukan proses iterasi. Berdasarkan pada gambar 3, dapat diketahui bahwa jumlah *cluster* yang diinginkan untuk proses analisis ini adalah sebanyak 3 kelompok. Pendeteksian seberapa banyak proses iterasi yang dilakukan dalam analisis *clustering* untuk 33 objek

pengamatan atau 33 provinsi, maka dapat dilihat pada tabel *iteration history*.

Iteration	Change in Cluster Centers		
	1	2	3
1	.791	1.463	.816
2	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 2. The minimum distance between initial centers is 4.161.

Gambar 4 banyaknya iterasi

Berdasarkan pada gambar 4 mengenai banyaknya iterasi yang dilakukan dalam analisis *clustering*, dapat diketahui terdapat dua kali proses iterasi yang dibutuhkan untuk mendapatkan pengelompokan objek pengamatan dengan tepat. Berdasarkan pada tabel *iteration history*, bahwa jarak minimum antar pusat *cluster* yang terjadi pada hasil proses iterasi adalah 1,463.

	Cluster		
	1	2	3
Zscore (Limbah_Keluarga)	.06035	1.85608	-.21294
Zscore(Limbah_Pabrik)	-.03819	2.79379	-.30618
Zscore(Limbah_Lainnya)	2.40747	1.23934	-.40520

Gambar 5 final cluster center

Pada tabel *final cluster centers* masih berkaitan dengan proses standarisasi atau proses normalisasi yang dilakukan pada tahap *pre-processing* untuk data pengamatan dengan menggunakan metode *z-score*. Nilai *z-score* untuk tiap-tiap variabel pada masing-masing *cluster* didapatkan dari nilai rata-rata untuk keseluruhan data pengamatan yang tergolong dalam *cluster* tersebut. Berdasarkan pada nilai *z-score* dari gambar 5, maka terdapat dua ketentuan yang berlaku, yakni:

- Nilai negatif dari setiap nilai *z-score* untuk variabel yang diteliti pada masing-masing *cluster* menandakan bahwa data pengamatan yang terdapat dalam *cluster* tersebut berada di bawah rata-rata populasi objek yang diteliti untuk variabel tersebut.
- Nilai positif dari setiap nilai *z-score* untuk variabel yang diteliti pada masing-masing *cluster* menandakan bahwa data pengamatan yang terdapat dalam *cluster* tersebut berada di atas rata-rata populasi objek yang diteliti untuk variabel tersebut.

Rumus umum yang digunakan, yaitu:

$$X = \mu + Z \cdot \sigma \quad (10)$$

X: rata-rata sampel (variabel dalam *cluster*)

μ : rata-rata populasi

Z: nilai standarisasi

σ : standar deviasi

Berdasarkan rumus umum tersebut, maka didapatkan bahwa rata-rata jumlah kasus pencemaran air yang berasal dari limbah keluarga pada *cluster* 1, *cluster* 2, dan *cluster* 3 secara berurutan adalah 83, 224, dan 62 kasus pencemaran air. Sedangkan rata-rata jumlah kasus pencemaran air yang berasal dari limbah pabrik pada setiap *cluster* secara berurutan adalah 65, 369, dan 37 kasus pencemaran air. Rata-rata selanjutnya yang dihitung dari jumlah kasus pencemaran yang berasal dari limbah lainnya selain dua limbah sebelumnya yang juga dijadikan sebagai variabel penelitian pada masing-masing *cluster* secara berurutan adalah 314, 203, dan 46 kasus pencemaran air.

Berdasarkan gambar 5 yang menunjukkan tabel *final cluster centers* dan dengan dua ketentuan yang telah disebutkan, maka hasil yang didapatkan dari tabel *final cluster centers* tersebut adalah:

a. Cluster -1 (Kelompok Pertama)

Dari tabel *final cluster centers*, maka diketahui bahwa pada *cluster*-1 atau kelompok pertama terdiri dari provinsi-provinsi di Indonesia yang memiliki jumlah kasus pencemaran air yang berasal dari limbah keluarga dan limbah lainnya (selain limbah keluarga dan pabrik) yang berada di atas rata-rata populasi provinsi yang diteliti.

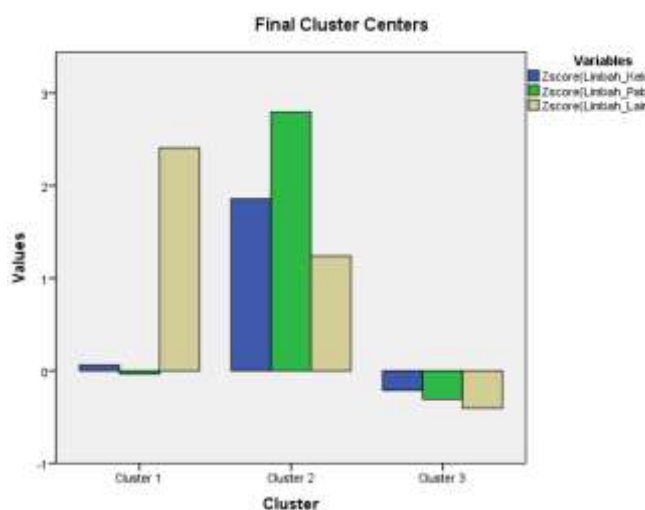
Sedangkan untuk instrumen lainnya yang diteliti, yakni limbah pabrik, dapat dikatakan bahwa provinsi-provinsi di Indonesia pada *cluster*-1 memiliki jumlah kasus pencemaran air yang berasal dari limbah pabrik berada di bawah rata-rata populasi provinsi yang diteliti. Oleh karena itu, dapat diduga bahwa pada kelompok pertama berisikan provinsi-provinsi yang menyumbangkan pencemaran air yang berasal dari daerah perkotaan, daerah pesisir pantai/laut, daerah pedesaan, dan daerah pegunungan.

b. Cluster -2 (kelompok Kedua).

Karakteristik provinsi yang digolongkan dalam *cluster*-2 merupakan provinsi-provinsi yang memiliki jumlah kasus pencemaran air dari limbah pabrik, limbah keluarga, dan limbah lainnya selain dari dua jenis limbah yang diteliti yang berada di atas rata-rata dari populasi provinsi yang diteliti. Oleh karena itu, dapat diperkirakan bahwa provinsi-provinsi yang berada pada kelompok kedua merupakan provinsi-provinsi yang menyumbangkan pencemaran air yang berasal dari daerah perkotaan, daerah pesisir pantai/laut, daerah pedesaan, daerah pegunungan, daerah industri.

c. *Cluster-3* (kelompok Ketiga)

Karakteristik provinsi yang digolongkan dalam *cluster-3* merupakan provinsi-provinsi yang memiliki jumlah kasus pencemaran air dari limbah pabrik, limbah keluarga, dan limbah lainnya selain dari dua jenis limbah yang diteliti yang berada di bawah rata-rata dari populasi provinsi yang diteliti. Oleh karena itu, dapat dikatakan bahwa provinsi yang terdapat pada *cluster-3* mengalami jumlah pencemaran air yang lebih sedikit dari ketiga jenis limbah yang diteliti dibandingkan dengan provinsi-provinsi lainnya.



Gambar 5 Bar chart of final cluster centers

Hal yang sama juga dapat dilihat dari *bar chart* untuk *final cluster centers*. Berdasarkan grafik tersebut, maka dapat diketahui bahwa kasus pencemaran terbanyak untuk *cluster-1* disebabkan oleh limbah lainnya, untuk *cluster-2* kasus pencemaran air terbanyak disebabkan oleh limbah pabrik. Sedangkan untuk *cluster-3* kasus pencemaran air terbanyak disebabkan oleh limbah keluarga, walaupun dari grafik menunjukkan bahwa kasus pencemaran air yang disebabkan oleh 3 jenis limbah yang diteliti berada di bawah rata-rata populasi provinsi di luar *cluster-3*.

Dari perhitungan rumus umum untuk rata-rata variabel yang diteliti dari masing-masing *cluster* dapat dikatakan sesuai atau valid berdasarkan pada hasil dari tabel *final cluster centers*.

Tabel 2 number of cases in each cluster

Cluster	1	3.000
	2	3.000
	3	27.000
Valid		33.000
Missing		.000

Dari Tabel 2 yang menunjukkan jumlah dari provinsi di setiap *cluster* adalah, pada *cluster* pertama didapatkan provinsi yang digolongkan dalam *cluster* ini sebanyak 3 provinsi, untuk *cluster* kedua terdapat 3 provinsi yang digolongkan dalam *cluster* kedua. Sedangkan pada *cluster* ketiga terdapat 27 provinsi yang terdapat di dalam *cluster* ini. Sehingga total keseluruhan provinsi yang diteliti adalah sebanyak 33 provinsi dan tidak ada objek pengamatan yang hilang selama proses analisis.

Case Number	Provinsi	Cluster	Distance
1	1.00	3	2.943
2	2.00	3	1.793
3	3.00	3	.774
4	4.00	3	.783
5	5.00	3	.798
6	6.00	3	2.207
7	7.00	3	.165
8	8.00	3	.484
9	9.00	3	1.119
10	10.00	3	.816
11	11.00	3	.741
12	12.00	2	1.463
13	13.00	2	.151
14	14.00	3	.743
15	15.00	2	1.442
16	16.00	3	.726
17	17.00	3	.368
18	18.00	3	.297
19	19.00	3	.646
20	20.00	1	.791
21	21.00	1	.269
22	22.00	1	.582
23	23.00	3	1.034
24	24.00	3	.429
25	25.00	3	.321
26	26.00	3	.201
27	27.00	3	.537
28	28.00	3	.614
29	29.00	3	.674
30	30.00	3	.633
31	31.00	3	.492
32	32.00	3	.676
33	33.00	3	.336

Gambar 6 cluster membership

Berdasarkan pada gambar 6 yang merupakan tabel *cluster membership*, dimana pada tabel tersebut menunjukkan pengelompokan untuk ke-33 provinsi di Indonesia dan jarak antara objek pengamatan, yakni 33 provinsi dengan pusat *cluster*. Oleh karena itu didapatkan hasil, sebagai berikut:

a. *Cluster-1* (Kelompok Pertama)

Pada *cluster-1* atau kelompok pertama, berisikan provinsi dengan nomor urut sesuai tabel 1, yakni provinsi nomor ke-20, 21, dan 22 atau provinsi Kalimantan Barat, Kalimantan Tengah, dan Kalimantan Selatan dengan masing-masing jarak dari provinsi-provinsi tersebut ke pusat *cluster-1* adalah 0,791; 0,269 ; dan 0,582.

b. *Cluster-2* (Kelompok Kedua)

Pada *cluster-2* atau kelompok kedua, berisikan provinsi dengan nomor urut sesuai tabel 1, yakni provinsi nomor ke-12, 13, dan 15 atau provinsi Jawa

Barat, Jawa Tengah, dan Jawa Timur dengan masing-masing jarak dari provinsi-provinsi tersebut ke pusat *cluster-2* adalah 1,463 ; 0,151 ; dan 1,442.

c. Cluster -3 (Kelompok Ketiga)

Pada cluster-3 kelompok tiga, berisikan provinsi dengan nomor urut sesuai pada gambar 6, yakni provinsi ke-1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 16, 17, 18, 19, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, dan 33 atau provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, NTB, NTT, Kalimantan timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, dan Papua dengan masing-masing jarak dari provinsi tersebut ke pusat cluster 3 adalah 2,934 ; 1,793 ; 0,774 ; 0,783 ; 0,798 ; 2,207 ; 0,165 ; 0,484 ; 1,119 ; 0,816 ; 0,741 ; 0,743 ; 0,726 ; 0,368 ; 0,297 ; 0,646 ; 1,034 ; 0,429 ; 0,321 ; 0,201 ; 0,537 ; 0,614 ; 0,674 ; 0,633 ; 0,492 ; 0,676 ; dan 0,336.

4. Kesimpulan dan Saran

A. Kesimpulan

Penerapan metode average linkage sebagai suatu metode penentuan jarak dan metode k-mean sebagai metode pengelompokan. Berikut adalah kesimpulan pada penelitian ini, antara lain:

1. Terdapat 28 provinsi pada kelompok pertama yakni dimana provinsi berada pada urutan 1-11, 14-19, dan 23-33. Pada kelompok kedua diisi oleh urutan 12 dan 13 sedangkan kelompok ketiga diisi oleh urutan 20 hingga urutan ke 22.
2. Sedangkan dalam variable limbah terdapat kelompok limbah keluarga dan limbah lainnya yang paling banyak mendominasi, kedua kelompok ini memiliki tingkat pencemaran yang paling tinggi.
3. Sehingga dapat disebut sebaran limbah di Indonesia populasinya masih merata di setiap provinsi, terlihat pada kelompok pertama dimana 28 provinsi tergolong sama. Ini menjelaskan bahwa tingkat persebaran limbah di Indonesia merata di setiap wilayah. Selanjutnya limbah keluarga menjadi yang terbesar terhadap pencemaran air.

B. Saran

Besarnya limbah limbah di Indonesia memiliki potensi besar terhadap keberlangsungan hidup masyarakat, sehingga dalam penelitian selanjutnya disarankan untuk meneliti terkait system klasifikasi produktivitas pengelolaan limbah di setiap provinsi. Sebab dalam penelitian ini limbah kerluarga menjadi limbah yang paling banyak memberikan dampak pencemaran terhadap perairan Indonesia.

Daftar Pustaka

- Aris, N. W., Ayu, I. G., & Sari, K., 2016, Pengelompokan Desa/Kelurahan di Kota Denpasar Menurut Indikator Pendidikan. E-Jurnal Matematika
- Santosa, B., 2007, Data Mining Terapan Dengan Matlab.: Graha Ilmu, Yogyakarta
- Amin, N., Kroeze, C., & Strokal, M., 2017, Human waste: An underestimated source of nutrient pollution in coastal seas of Bangladesh, India and Pakistan. Marine Pollution Bulletin
- Chevez, P., Barbero, D., Martini, I., & Discoli, C., 2017, Application of the k-means clustering method for the detection and analysis of areas of homogeneous residential electricity consumption at the Great La Plata region, Buenos Aires, Argentina. Journal Sustainable Cities and Society
- Kapageridis, i. K., 2015, Variable lag variography using k-means clustering. Journal Computers and Geosciences
- Panjaitan, M., 2000, Memberdayakan Kaum Miskin, BPK Gunung Mulia, Jakarta
- Wahyono, S., 2009, Indonesia Negara Maritim, Ikatan Penerbit Indonesia, Jakarta