

Perbandingan Pengelompokan pada Wilayah Berdasarkan Tingkat Kerusakan Lingkungan di Indonesia

Muhammad Rianbarenito Gunawan¹, Yaya Setiadi²

Badan Pusat Statistik¹
Politeknik Statistika STIS²
setiadi@stis.ac.id

Abstrak—Kerusakan lingkungan adalah salah satu perhatian dunia saat ini. Dikarenakan kerusakan lingkungan akan memiliki dampak negatif bagi umat manusia. Oleh karena itu, untuk memberikan gambaran mengenai daerah yang butuh perhatian lebih dalam masalah kerusakan lingkungan di Indonesia, diperlukan pengelompokan. Penelitian ini menggunakan metode hierarki, *k-means*, kluster, dan kluster *ensemble* untuk mengelompokkan wilayah. Hasil analisis menunjukkan bahwa kluster *ensemble* memberikan hasil yang lebih baik dalam pengelompokan wilayah. Pada penelitian ini digunakan data dari 34 provinsi dengan menggunakan variabel indeks kualitas udara (IKU), indeks kualitas air (IKA), kualitas tutupan lahan (IKTL), dan sampah terkelola.

Kata kunci: *Lingkungan, analisis kluster, kluster ensemble, kluster hierarki*

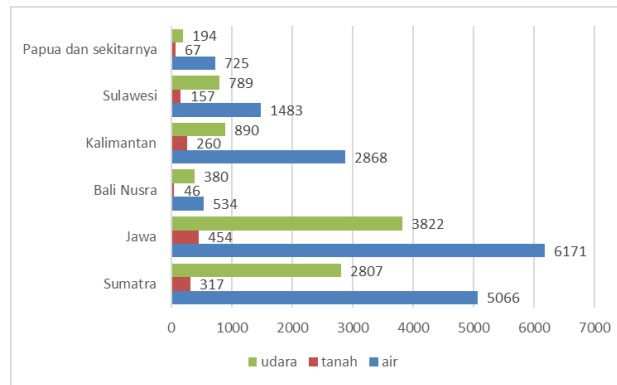
I. PENDAHULUAN

Degradasi lingkungan adalah kerusakan terhadap lingkungan melalui penipisan sumber daya seperti udara, air dan tanah, juga kerusakan ekosistem, habitat dan kepunahan kehidupan liar serta polusi. Hal ini didefinisikan juga sebagai perubahan atau gangguan terhadap lingkungan yang merusak atau tidak diinginkan. Masalah lingkungan dapat mengancam kehidupan manusia, pada tahun 2004, High Level Threat Panel, Challenges and Change PBB, memasukkan degradasi lingkungan sebagai salah satu dari sepuluh ancaman terhadap kemanusiaan. Sejak tahun 2000, Bank Dunia telah melakukan upaya sistematis untuk mengukur kerusakan lingkungan di tingkat nasional dan lokal di beberapa negara. Cost of Environmental Degradation (COED) yang dilakukan bank dunia biasanya mengukur kerusakan yang disebabkan oleh beberapa kategori lingkungan: air, udara, lahan pertanian, hutan, limbah, dan zona pesisir seperti pada [15].

Dilihat dari sisi kesehatan, kerusakan lingkungan memberikan dampak yang serius. Pencemaran tanah, udara dan air saja menyebabkan sembilan juta kematian pada tahun 2015 (Landrigan et al., 2018). Penelitian oleh Lelieveld et al (2019) memperkirakan 8,8 juta kematian akibat polusi udara dalam dan luar ruangan secara total pada tahun 2015. Menurutnya 5,5 juta orang meninggal sebelum waktunya setiap tahun karena polusi udara dari semua sumber antropogenik. Selain itu, kerusakan lingkungan memberikan dampak serius bagi ekonomi dan sosial. Sebagai contoh dengan mempertimbangkan hanya kematian dini, polusi udara merugikan ekonomi dunia sekitar US\$225 miliar dalam bentuk kehilangan pendapatan tenaga kerja dan US\$5 triliun dalam kerugian kesejahteraan (Bank Dunia dan Institute for Health Metrics and Evaluation, 2016).

Sebagai akibat dari perubahan iklim dan bentuk degradasi lingkungan, terjadi peningkatan frekuensi dan intensitas kejadian cuaca ekstrim dan bencana. Dengan adanya bencana tersebut akan timbul kerugian, berdasarkan ILO (2018) menunjukkan bahwa antara tahun 2000 dan 2015, 23 juta tahun kehidupan kerja hilang setiap tahun sebagai akibat dari berbagai bahaya terkait lingkungan yang disebabkan atau ditingkatkan oleh aktivitas manusia. Dimana hal ini setara dengan 0,8 persen pekerja setahun, mengingat 2,8 miliar orang berusia 15 hingga 64 tahun bekerja pada tahun tertentu. Selain itu, menurut Montt (2018) masalah lingkungan seperti polusi udara dengan demikian meningkatkan ketidaksetaraan gender di pasar tenaga kerja.

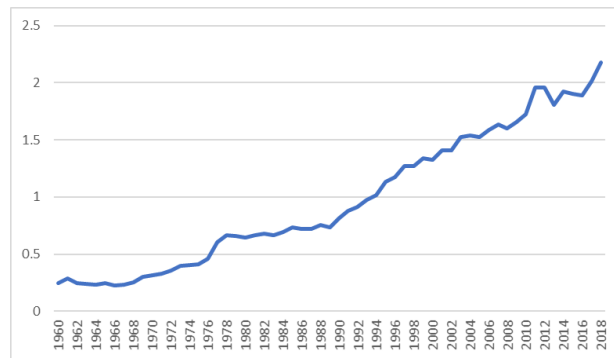
Berkaca pada negara Indonesia, masih cukup banyak daerah yang mengalami permasalahan lingkungan. Berdasarkan data Podes pada tahun 2018, dari 83.931 desa/kelurahan sekitar 26 persen desa mengalami permasalahan lingkungan (BPS 2018).



Sumber : Podes, 2018

GAMBAR 1. BANYAKNYA DESA/KELURAHAN YANG MENGALAMI MASALAH LINGKUNGAN.

Dapat dilihat pada gambar bahwa pencemaran udara dan air cukup mendominasi di Indonesia. Pulau Sumatra dan Jawa adalah daerah dengan desa/kelurahan yang paling banyak mengalami masalah lingkungan. Ada 3.822 dan 6.171 desa/kelurahan yang mengalami masalah pencemaran udara dan air di Pulau Jawa. Pada Pulau Sumatra ada sekitar 27,26 persen desa/kelurahan yang memiliki masalah lingkungan, dimana sebagian besar masalah lingkungan ini berkaitan dengan pencemaran udara dan air. Hal berbeda terjadi pada Pulau Kalimantan dimana ada 2.868 desa/kelurahan yang mengalami pencemaran air atau sekitar 39,6 persen dibandingkan seluruh desa/kelurahan yang ada di Kalimantan. Pada daerah lain juga memiliki masalah lingkungan namun secara rasio tidak sebanyak daerah yang telah disebutkan tadi.



Sumber : worldbank, 2021

GAMBAR 2. EMISI CO₂ PER KAPITA DI INDONESIA (METRIK TON) DARI TAHUN 1960 HINGGA 2018

Jika dilihat dalam rentang waktu yang panjang kualitas lingkungan di Indonesia semakin menurun. Referensi [13] mengatakan bahwa emisi karbon merupakan salah satu ukuran lingkungan dikarenakan semakin tinggi emisi karbon maka akan merusak lingkungan dan meningkatkan potensi perubahan iklim. Di Indonesia terjadi kecenderungan kenaikan emisi karbon. Secara rata-rata tiap individu mengalami kenaikan kenaikan emisi CO₂ sebesar 0,03 metrik setiap tahun. Pada tahun 2018 emisi yang dihasilkan berada pada angka 2,17 metrik ton, hal ini jauh dibandingkan pada tahun 1960 yang hanya 0,24 metrik ton. Beberapa dampak yang timbul akibat adanya efek rumah kaca adalah perubahan temperatur bumi yang semakin tinggi sehingga menyebabkan perubahan iklim. Indonesia sebagai negara yang berada di ekuator bumi tentu akan sangat terdampak buruk karena perubahan iklim.

Pencegahan kerusakan lingkungan dengan cara pembangunan berkelanjutan sangatlah penting, yaitu pembangunan yang memperhatikan lingkungan dan melestarikan fungsi ekosistem. Diperlukan juga dukungan masyarakat luas dan pemerintah yang berperan aktif dan kebiasaan hidup yang memperhatikan kelestarian lingkungan dan kemampuan ekosistem, baik masa kini maupun yang akan datang. Namun sayangnya kerusakan lingkungan sudah terjadi di Indonesia dan perlu tindakan yang serius akan hal ini.

Oleh karena itu, guna memberikan gambaran mengenai daerah yang butuh perhatian lebih dalam masalah kerusakan lingkungan perlu dilakukan pengelompokan tingkat kerusakan lingkungan di Indonesia.

Pengelompokan daerah berdasarkan tingkat kerusakan lingkungan dapat dilakukan dengan *clustering*. Menariknya, berbedanya metode analisis klaster yang digunakan akan berakibat pula pada kemungkinan terjadinya perbedaan hasil pengelompokan. Oleh karena itu, membandingkan hasil dari beberapa metode pengelompokan menjadi suatu yang penting dan lebih bermanfaat. Banyak metode *clustering* yang telah dikembangkan oleh para pakar dan telah banyak pula diterapkan pada berbagai bidang. Selain metode konvensional, yaitu analisis klaster berhirarki dan *non*-hirarki, salah satu metode yang hingga saat ini banyak dikembangkan adalah metode klaster ensemble. Metode ini diperkenalkan oleh Strehl dan Gosh (2002). Ide dasar dari klaster ensemble adalah mengombinasikan sekumpulan hasil gerombol dari metode-metode lain. Menurut penelitian Strehl dan Gosh tersebut, klaster *ensemble* dapat memberikan hasil penggerombolan yang lebih berkualitas. Berdasarkan uraian tersebut, penelitian ini bertujuan untuk membandingkan pengelompokan pada wilayah berdasarkan tingkat kerusakan lingkungan di Indonesia menggunakan analisis klaster hirarki, k-means, dan ensemble. Sehingga dengan begitu diharapkan Mengetahui hasil pengelompokan tingkat kerusakan lingkungan pada provinsi-provinsi yang ada di Indonesia. Lalu dapat membandingkan metode mana yang paling tepat diantara analisis klaster hirarki, K-means, dan ensemble.

II. METODE PENELITIAN

2.1 Landasan Teori

Lingkungan merupakan tempat hidup sekaligus menjadi tempat penampungan limbah hasil aktivitas manusia. Lingkungan memiliki kemampuan bertahan dalam keadaannya dan menetralkan diri kembali ke keadaan awal jika limbah tersebut masih berada dalam batas daya dukung lingkungan tersebut (sumampouw dan risjani, 2018, hal 1). Oleh karena itu, diperlukan ukuran dalam menilai kerusakan lingkungan. Salah satu ukuran untuk melihat kerusakan lingkungan dilakukangan oleh Bank Dunia. *Metode Cost of Environmental Degradation* (COED) yang dilakukan bank dunia biasanya mengukur kerusakan yang disebabkan oleh beberapa kategori lingkungan yaitu: air, udara, lahan pertanian, hutan, limbah, dan zona pesisir. Di Indonesia sendiri kementerian lingkungan hidup dan kehutanan (KLHK) melakukan pengukuran lingkungan di 34 provinsi di Indonesia. Berikut adalah data yang dikumpulkan oleh KLHK :

a) Indeks kualitas udara

Indeks kualitas udara (IKU) adalah ukuran yang digunakan untuk menilai kualitas udara. Indeks ini biasa digunakan untuk memperlihatkan seberapa buruk kualitas udara di suatu daerah. Pemantauan kualitas udara ambien pada umumnya mencakup nilai beberapa parameter pencemar udara seperti SO₂, NO₂, O₃, CO, PM_{2.5}, PM₁₀, dll.

b) Indeks kualitas air

Indeks kualitas udara (IKA) adalah ukuran yang digunakan untuk menilai kualitas air. Berdasarkan Keputusan Menteri Lingkungan Hidup Nomor 115 Tahun 2003, bahwa metode untuk menentukan kualitas air dapat digunakan metode storet maupun metode indeks pencemaran. Dalam hal ini Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia (KLHK) menggunakan metode indeks pencemaran, sehingga dengan begitu Indeks pencemaran air dapat digunakan untuk menilai kualitas badan air, dan kesesuaian peruntukan badan air tersebut.

c) Sampah terkelola

Menurut Departemen PU (2004) timbulan sampah adalah volume sampah atau berat sampah yang dihasilkan dari jenis sumber sampah di wilayah tertentu persatuan waktu. KLHK sendiri mendefinisikan sampah terkelola adalah pengurangan sampah ditambah sampah yang berhasil dikelola pada wilayah dan periode tertentu lalu dibagi dengan jumlah timbulan sampah pada wilayah dan pada periode waktu yang sama juga.

d) Indeks Kualitas tutupan lahan

Indeks Kualitas tutupan lahan (IKTL) didapatkan perbandingan antara tutupan lahan berhutan dengan luas wilayah provinsi. Berdasarkan UU Nomor 41 Tahun 1999, bahwa setiap Provinsi minimal memiliki kawasan hutan sekitar 30 persen dari luas wilayah.

2.2 Cakupan Penelitian

Pada penelitian ini digunakan empat variabel untuk pengelompokan data dari 34 provinsi yaitu indeks kualitas udara (IKU), indeks kualitas air (IKA), kualitas tutupan lahan (IKTL), dan sampah terkelola. Data yang digunakan adalah data tahun 2019 yang didapatkan dari data sekunder KLHK. Pemilihan data tahun 2019 dilakukan agar lebih merepresentasikan kerusakan lingkungan. Hal ini dikarenakan pada tahun 2020 dan 2021 terjadi pandemi sehingga seakan-akan aktivitas yang menimbulkan kerusakan lingkungan menurun.

2.3 Metode Analisis

Penelitian ini merupakan jenis penelitian kluster. Analisis kluster merupakan salah satu teknik Multivariate yang masuk dalam klasifikasi metode interdependen. Analisis kluster bertujuan untuk mengelompokkan objek atau kasus ke dalam kelompok yang mempunyai sifat yang relatif homogen. Sebelum melangkah pada metode pengklasteran, dilakukan preparasi data dahulu seperti imputasi dan standarisasi pada data untuk menyamakan satuan. Dalam pengklasteran, pengukuran jarak yang digunakan dalam penelitian ini adalah *Euclidean distance*, yang dalam hal ini berguna untuk mengukur kemiripan atau ketidakmiripan antar objek.

Imputasi regresi

Ada beberapa teknik mengatasi missing value pada data salah satunya imputasi regresi. Regresi Linier adalah model yang didasari oleh hubungan antara dua variabel dengan menyesuaikan persamaan linier dengan yang diamati pada data. Nilai data yang hilang akan diganti dengan regresi variabel yang tidak teramati terhadap yang diamati pada data (Zakaria & Noor, 2018). Persamaan direpresentasikan sebagai berikut :

$$y_i^* = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (1)$$

beberapa uji asumsi klasik yang harus dipenuhi untuk menggunakan model regresi, yaitu : uji autokorelasi, uji heteroskedastisitas, dan uji normalitas.

Analisis kluster K-means

Algoritma K-Means berdasarkan pembagian adalah sejenis algoritma kluster yang diusulkan oleh J. B. MacQueen. Algoritma yang unsupervised ini biasanya digunakan dalam data mining dan melihat pola. Dasar dari algoritma ini bertujuan untuk meminimalkan indeks kinerja kluster, kesalahan kuadrat, dan kesalahan criterion. Dalam penggunaan jumlah titik inisiasi yang berbeda maka akan menghasilkan hasil yang berbeda pula dalam metode ini. Lebih lanjut metode ini pengelompokan dilakukan terlebih dahulu memilih satu objek dasar yang akan dijadikan nilai awal kluster, kemudian semua kluster yang ada dalam jarak terdekat dengan kluster ini akan bergabung, demikian seterusnya sehingga terbentuk beberapa kluster dengan keseluruhan objek yang terdapat didalamnya. Proses penentuan centroid dan penempatan objek dalam kluster diulangi sampai nilai centroid konvergen (centroid dari semua kluster tidak berubah lagi).

Analisis kluster Hirarki

Metode berhirarki digunakan untuk mengelompokkan objek secara terstruktur berdasarkan kemiripan sifatnya dan kluster yang diinginkan belum diketahui banyaknya seperti pada [10]. Terdapat dua prosedur pada metode berhirarki, yaitu prosedur aglomeratif dan prosedur divisive seperti pada [6]. Sedangkan beberapa ukuran kemiripan dan ketidakmiripan antar kluster dapat dilihat dengan menggunakan Pautan Tunggal, Pautan Lengkap, Pautan Centroid, Pautan Rataan, Pautan Median dan Ward. Hasil pengklasteran dengan metode berhirarki dapat digambarkan dalam sebuah diagram pohon yang biasa disebut dendrogram Menurut Mattjik & Sumertajaya (2011), banyaknya kluster yang terbentuk ditentukan dari dendrogram yang terjadi dan tergantung subjektivitas peneliti. Namun demikian pemisahan kluster biasanya ditentukan berdasarkan jarak penggabungan terbesar.

DIANA

Divisive Analysis Clustering (DIANA) adalah teknik pengelompokan secara hirarki yang mana membangun hierarkinya dalam urutan terbalik sehingga dikenal juga sebagai pendekatan *top-down*. Metode Ini membuat kelompok dengan membagi data awal, yang mana merupakan kebalikan dari algoritma

Agglomerative Hierarchical Clustering dimana data dikelompokkan secara aglomeratif. Pada metode ini, awalnya seluruh objek dikelompokkan pada cluster yang sama, dan dilanjutkan dengan pemisahan cluster secara rekursif sampai data individu telah dipecah menjadi cluster tunggal.

PAM

Algoritma *k-medoids* atau dikenal juga *partition around medoids* (PAM) adalah salah satu teknik pengklasteran yang mirip dengan *k-means* dimana berguna dalam analisis pengelompokan. Berbeda dengan *k-means* yang menggunakan rata-rata, PAM mengelompokkan data/objek berdasarkan *k-medoid* yang dipilih sebagai pusat kluster. Metode ini menggunakan data yang terletak di tengah cluster, maka metode ini lebih kuat terhadap outlier dibandingkan dengan metode *k-means* (Kaufman & Rousseuw 1990). Dalam hal ini PAM meminimalkan perbedaan data dengan data terdekatnya sehingga dapat terbentuk kelompok.

Analisis Kluster Ensemble

Cluster Ensemble diperkenalkan oleh Strehl dan Gosh (2002), yaitu sebuah metode yang digunakan untuk mengkombinasikan sekumpulan solusi kluster. Metode ini memiliki keunggulan dibanding metode pengklasteran lainnya, yakni mampu meningkatkan kualitas dan kekekaran solusi kluster. Tantangan untuk mendapatkan solusi cluster dengan kualitas yang baik dan adanya keragaman solusi cluster yang dihasilkan dari metode yang berbeda merupakan motivasi dikembangkannya Kluster Ensemble. Pengklasteran pada Kluster Ensemble dilakukan dengan mengkombinasikan berbagai solusi dari berbagai metode pengklasteran hingga diperoleh satu pengklasteran akhir yang lebih baik. Input yang dibutuhkan adalah solusi pengklasteran yang telah diperoleh dengan menggunakan berbagai hasil pengklasteran tanpa melihat karakteristik data awal. Menurut Chiu & Talhouk (2018) *ensemble clustering* dilakukan dalam beberapa tahap. Dimana n = sampel, p = fitur a = algoritma, dan k = banyak kluster.

- Pada tahap pertama data dari beberapa variabel yang digunakan akan dibentuk dalam matrix yang mana terdiri kolom n dan baris p , data matrix ini akan digunakan sebagai input pada proses selanjutnya.
- Selanjutnya pada tahap kedua dijalankan sebanyak A algoritma dengan K sebagai jumlah kluster yang diinginkan. Pada tahap ini data diaplikasikan beberapa algoritma misal *k-means*, hirarki, PAM, dan lain-lain sehingga akan didapatkan hasilnya.
- Untuk tahap ketiga dilakukan consensus functions yaitu penggabungan kluster sebanyak A algoritma. Ada beberapa consensus functions yang dapat digunakan misalnya majority voting, *K-modes*, LCE, dan lain-lain.
- Dengan demikian, terakhir didapatkan hasil konsensus diri seluruh sampel dan algoritma.

III. HASIL DAN PEMBAHASAN

Evaluasi Hasil Klustering

Selanjutnya evaluasi dilakukan untuk menilai kualitas kluster yang dihasilkan. Ada berbagai jenis ukuran yang dapat digunakan. Pada [3] juga membagi evaluasi kluster menjadi dua kategori, dimana *compactness* adalah sebuah ukuran untuk melihat seberapa erat hubungan objek dalam kluster. Sedangkan *separation* mengukur seberapa berbeda atau terpisah antara kluster dengan kluster lainnya. Sehingga berdasarkan hal tersebut pada penelitian ini digunakan dua metode untuk evaluasi kluster yaitu *compactness* dan *davis bouldin*.

- Compactness (CP)

Compactness (CP) merupakan salah satu pengukuran ketepatan pengklasteran yang umum digunakan. CP mengukur jarak rata-rata antara setiap pasang titik data yang termasuk dalam kelompok yang sama

$$CP = \frac{1}{N} \sum_{k=1}^k n_k \left(\frac{\sum_{x_i, x_j \in C_k} d(x_i, x_j)}{n_k(n_k-1)/2} \right) \quad (2)$$

dengan K adalah banyaknya cluster yang terbentuk, n_k adalah jumlah data yang termasuk ke dalam cluster ke- K , $d(x_i, x_j)$ adalah jarak antara objek ke- i dan objek ke- j , dan n adalah jumlah seluruh objek. Hasil pengklasteran dikatakan lebih baik jika Semakin kecil nilai indeks Compactness.

b) Davies-Bouldin (DB)

Indeks DB merupakan salah satu metode evaluasi internal yang mengukur evaluasi cluster pada suatu metode pengklasteran yang didasarkan pada nilai kohesi dan separasi. Indeks DB dirumuskan sebagai berikut:

$$DB = \frac{1}{k} \sum_{k=1}^k R_i \quad (3)$$

$$\text{Dengan } R_i = \max_{j=1, \dots, k, i \neq j} R_{ij}$$

$$R_{ij} = \frac{S_i + S_j}{d_{ij}} \quad (4)$$

Dengan $d_{ij} = d(x_i, x_j)$. Dimana V_i adalah nilai rata-rata pada cluster C_i dan V_j adalah nilai rata-rata pada cluster C_j

$$S_i = \frac{1}{|C_i|} \sum_{v \in C_i} d(x, v_i) \quad (5)$$

dengan $|C_i|$ adalah jumlah data yang termasuk ke dalam cluster ke-i. Skema pengklasteran Davies Bouldin yang optimal adalah yang memiliki nilai paling kecil. Semakin kecil nilai pada indeks, maka hasil pengklasteran semakin baik.

Pada penelitian ini digunakan empat variabel untuk pengelompokan data dari 34 provinsi ke kluster tertentu yaitu indeks kualitas udara (IKU), indeks kualitas air (IKA), kualitas tutupan lahan (IKTL), dan sampah terkelola. Ada beberapa preparasi data yang dilakukan guna mengatasi permasalahan data. Pada data indeks kualitas tutupan lahan pada tahun 2019 provinsi Kalimantan utara dan Kalimantan timur belum terpisah menjadi provinsi masing-masing. Untuk mengatasi hal tersebut untuk provinsi Kalimantan utara dan Kalimantan timur digunakan nilai IKTL yang sama, dikarenakan indeks tersebut adalah nilai rata-rata dari gabungan kedua provinsi tersebut. Selanjutnya pada variabel sampah terkelola pada provinsi Nusa Tenggara Timur pada tahun 2019 adalah 0 persen. Hal ini terjadi karena belum adanya basis pengumpulan data maka dilakukan imputasi regresi agar datanya tidak kosong yaitu dengan bantuan data 2020 untuk kasus ini.

Persamaan model regresi linier yang dibentuk dengan menggunakan regresi linier

$$y_i^* = 1,312 + 0,007X_i \quad (6)$$

y_i^* : sampah terkelola provinsi ke-I pada tahun 2019 (nilainya ditransformasi dalam bentuk Log)

X_i : sampah terkelola provinsi ke-I pada tahun 2020

Sebelum dilakukan regresi data pada sampah terkelola pada tahun 2019 dilakukan transformasi dalam bentuk Logaritma basis 10 guna mengatasi ketidaknormalan. Selanjutnya dilakukan pengujian asumsi klasik yaitu uji normalitas dan uji homoskedastis. Berdasarkan uji *Kolmogorov Smirnov* dan mendapatkan p-value sebesar 0,075, nilai tersebut lebih besar dari alpha yaitu 0,05, artinya residual berdistribusi normal. Pengujian asumsi homoskedastis dilakukan dengan uji *glejser* dimana nilai p-value adalah 0.149 dimana lebih besar dari alpha yaitu 0,05 sehingga dapat disimpulkan bahwa tidak terjadi heteroskedastisitas. Nilai dari pengujian *Durbin-Watson* adalah 2,249, jika dibandingkan dengan tabel untuk $k=1$ dan $n=33$ didapatkan tidak terjadi pelanggaran asumsi. Hasil dari model didapatkan nilai *r square* adalah 0,568. Selanjutnya pada pengujian signifikansi parameter dengan alpha sebesar 5 persen didapatkan bahwa sampah terkelola pada tahun 2020 signifikan terhadap sampah terkelola pada tahun 2019. Pada persamaan regresi didapatkan nilai sampah terkelola pada provinsi NTT pada tahun 2019 adalah sebesar 33,11 persen.

TABEL 1. STATISTIK DESKRIPTIF

Variabel	Minimum	Maksimum	Jangkauan	Rata-rata	Standar deviasi	Varian
(1)	(2)	(3)	(4)	(5)	(6)	(7)
IKU	67,97	93,79	25,82	87,925	5,527	30,549
IKA	35,37	69,29	33,92	53,545	7,468	55,783
IKTL	24,66	100,00	75,34	61,888	19,992	399,684
Sampah terkelola	26,25	98,19	71,94	56,518	20,183	407,362

Data indeks kualitas udara (IKU) untuk 34 provinsi secara rata-rata memiliki nilai 87,925 yang mana nilai ini adalah nilai rata-rata nasional. Pada data ini terdapat 22 provinsi memiliki nilai diatas rata-rata yang artinya pada daerah tersebut memiliki kualitas udara yang lebih baik dibandingkan rata-rata nasional. Begitu pula sebaliknya terdapat 12 provinsi memiliki kualitas udara di bawah rata-rata nasional. Pada IKU terburuk dengan nilai 67,97 dan 74,98 terdapat pada Provinsi Jakarta dan Banten. Selanjutnya provinsi dengan nilai IKU terdapat pada Provinsi Kalimantan Utara dengan nilai 93,79. Pada variabel ini tidak terdapat pencilan ekstrim.

Data indeks kualitas air (IKA) untuk 34 provinsi secara rata-rata memiliki nilai 53,545 yang mana nilai ini adalah nilai rata-rata nasional. Pada data ini terdapat 19 provinsi memiliki nilai diatas rata-rata yang artinya pada daerah tersebut memiliki kualitas air yang lebih baik dibandingkan rata-rata nasional. Begitu pula sebaliknya terdapat 15 provinsi memiliki kualitas air dibawah rata-rata nasional. Pada IKA terburuk dengan nilai 35,37 terdapat pada Provinsi DI Yogyakarta. Selanjutnya provinsi dengan nilai IKA terbaik terdapat pada Provinsi Bangka Belitung dengan nilai 93,79. Pada variabel ini tidak terdapat pencilan ekstrim.

Data indeks kualitas tutupan lahan (IKTL) untuk 34 provinsi secara rata-rata memiliki nilai 61,888 yang mana nilai ini adalah nilai rata-rata nasional. Pada data ini terdapat 15 provinsi memiliki nilai diatas rata-rata yang artinya pada daerah tersebut memiliki kualitas tutupan lahan yang lebih baik dibandingkan rata-rata nasional. Begitu pula sebaliknya terdapat 19 provinsi memiliki kualitas tutupan lahan dibawah rata-rata nasional. Pada IKTL terburuk dengan nilai 24,66 terdapat pada Provinsi DKI Jakarta. Selanjutnya provinsi dengan nilai IKTL terbaik terdapat pada Provinsi Papua Barat dengan nilai 100. Pada variabel ini tidak terdapat pencilan ekstrim.

Sampah terkelola untuk 34 provinsi secara rata-rata memiliki nilai 56,51 persen yang mana nilai ini adalah nilai rata-rata nasional. Pada data ini terdapat 14 provinsi memiliki nilai diatas rata-rata yang artinya pada daerah tersebut memiliki Pengelolaan sampah yang lebih baik dibandingkan rata-rata nasional. Begitu pula sebaliknya terdapat 20 provinsi memiliki pengelolaan sampah dibawah rata-rata nasional. Pada sampah terkelola terburuk dengan nilai 26.25 persen terdapat pada Provinsi Sulawesi Tenggara. Selanjutnya provinsi dengan nilai Sampah terkelola terbaik terdapat pada Provinsi Papua dengan nilai 98,19 persen. Pada variabel ini tidak terdapat pencilan ekstrim.

Pengecekan Multikolinearitas

Langkah pertama dalam pengelompokan metode k-means adalah dengan melakukan pengecekan multikolinearitas untuk menentukan jarak yang tepat. Multikolinearitas berhubungan dengan situasi dimana ada adanya hubungan linier yang sempurna atau tepat di antara beberapa atau semua variabel penjelas dari model regresi (Gujarati, 2003). Sehingga dalam hal ini multikolinearitas mengacu pada keberadaan lebih dari satu linear yang tepat hubungan. Adapun hasil pengecekan multikolinearitas disajikan dalam tabel sebagai berikut.

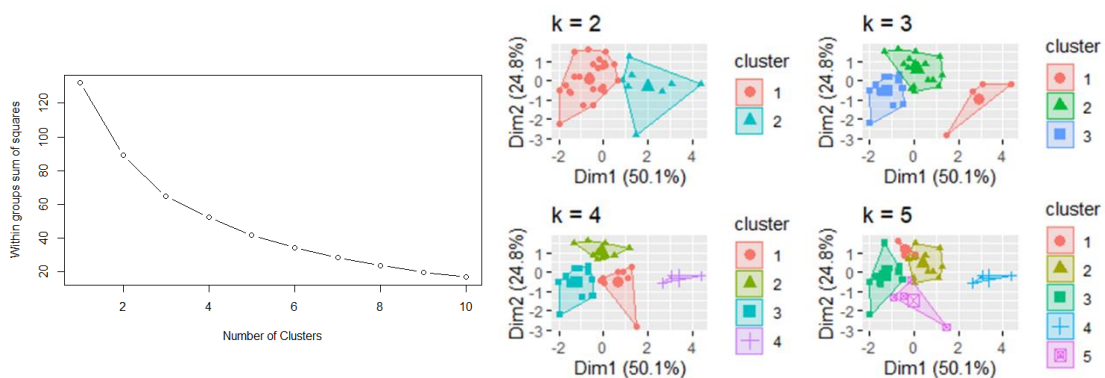
Tabel 2. Pengecekan Multikolinearitas pada variabel

Variabel	Tolerance	VIF
(1)	(2)	(3)
IKU	0.255	3.923
IKA	0.743	1.345
IKTL	0.866	1.154
Sampah terkelola	0.586	1.706

Berdasarkan hasil pengujian multikolinearitas pada tabel, nilai VIF dari variabel IKU, IKA, IKTL, dan Sampah terkelola memiliki nilai VIF kurang dari 10 dan nilai tolerance diatas 0.1. Dapat disimpulkan syarat non-multikolinearitas dalam menggunakan jarak euclidean terpenuhi.

Metode K-means

Pada pengklasteran yang pertama digunakan metode k-means. Penggunaan metode ini dilakukan karena metode k-means adalah metode yang paling rendah tingkat kompleksitasnya dibandingkan metode PAM, DIANA, Hirarki, maupun *ensemble*. Disamping itu penggunaan k-means sebagai pendeteksi jumlah kluster lebih mudah diinterpretasikan dibandingkan metode yang telah disebutkan sebelumnya. Langkah pertama yang dilakukan standarisasi pada data guna menyamakan satuan yang digunakan. Lalu dengan bantuan metode K-means dilakukan perhitungan *total within sum of square* dari jumlah kluster satu hingga ke-n sehingga didapatkan gambar elbow seperti pada gambar dibawah ini.



GAMBAR 3. (A) BAGIAN KIRI, HASIL PERHITUNGAN *TOTAL WITHIN SUM OF SQUARE* DALAM BENTUK GRAFIK *ELBOW*. (B) BAGIAN KANAN, PCA DARI VARIABEL YANG DIGUNAKAN UNTUK KLUSTER K=2 HINGGA K=5

Dengan melihat gambar, dapat dilihat bahwa penurunan *total within sum of square* yang tajam terjadi pada kluster dua dan berlanjut pada kluster tiga dan empat. Pada kluster yang berjumlah dua terjadi penurunan *sum of square* yang tertinggi yaitu sebesar 35 persen jika dibandingkan dengan kluster yang berjumlah satu. Secara grafik Dapat dilihat *elbow* atau kurva penurunan yang curam terjadi di sekitar kluster k = 2 Untuk kluster k = 5 keatas kurva sudah terlihat melandai dan tidak curam. Untuk menambah keyakinan mengenai pemilihan inisiasi jumlah kluster maka dilakukan *principal component analysis* (PCA), dapat dilihat bahwa jumlah kluster k=3 hingga k=4 sudah cukup baik karena data terbagi berdasarkan klasternya dan tidak saling tumpang tindih. Sehingga berdasarkan pertimbangan diatas, dalam penelitian ini peneliti menggunakan jumlah kelas adalah tiga sehingga nantinya diperoleh tingkat kerusakan lingkungan rendah, sedang, dan tinggi.

TABEL 3. RATA-RATA VARIABEL BERDASARKAN PENGLASTERAN K-MEANS

Tingkat kerusakan lingkungan	IKU	IKA	IKTL	Sampah terkelola	Rata-rata
(1)	(2)	(3)	(4)	(5)	(6)
Tinggi	75,81	41,50	33,80	52,98	51,02
Sedang	88,62	54,92	53,86	47,00	61,10
Rendah	90,74	55,46	81,03	70,06	74,32

Berdasarkan metode K-means pada kluster pertama hanya terdapat pada pulau jawa yang mana ada provinsi DKI Jakarta, Banten, Jawa barat dan DI Yogyakarta. Nilai rata-rata untuk semua indikator kerusakan lingkungan pada kluster pertama lebih rendah dibanding dengan nilai rata-rata provinsi di Indonesia dan juga paling rendah bila dibandingkan dengan rata-rata cluster lainnya. Dimana pada kelompok ini rata-rata gabungan dari IKU, IKA, IKTL, dan Sampah terkelola sebesar 51,02. Oleh karena itu, pada kluster pertama dapat dikatakan sebagai cluster yang memiliki tingkat kerusakan lingkungan yang paling tinggi dibanding cluster kedua dan ketiga.

Selanjutnya, pada pengelompokan kluster kedua terdiri dari provinsi yang cukup menyebar di berbagai pulau di Indonesia kecuali provinsi yang terletak di wilayah timur. Pada kelompok ini terdiri dari 17 provinsi. Daerah yang termasuk dalam kelompok ini memiliki nilai rata-rata gabungan IKU, IKA, IKTL dan Sampah Terkelola yaitu 61,10. kluster kedua dapat dikatakan kluster yang memiliki karakteristik kerusakan lingkungan sedang karena nilai rata-rata semua indikator kerusakan lingkungan di cluster kedua masih relatif tidak terlalu jauh dari nilai rata-rata Indonesia. Pada kelompok ini karakteristiknya relatif berada di tengah-tengah jika dibandingkan pada kelompok tingkat kerusakan lingkungan tinggi dan rendah.

Untuk yang terakhir yaitu pada pengelompokan kluster ketiga, dimana terdiri dari provinsi yang secara geografis terletak di wilayah timur Indonesia ataupun berada di wilayah Kalimantan pada kluster ini terdapat 13 provinsi. Daerah yang termasuk dalam kelompok ini memiliki nilai rata-rata gabungan IKU, IKA, IKTL dan Sampah Terkelola yaitu 74,32. kluster ketiga dapat dikatakan kluster yang memiliki tingkat kerusakan lingkungan rendah karena rata-rata nilai karakteristik lebih tinggi dibandingkan kluster satu dan dua.

Metode Hirarki

Selanjutnya pengelompokan provinsi-provinsi di Indonesia berdasarkan kerusakan lingkungan dilakukan menggunakan analisis kluster berhirarki. Banyaknya cluster yang terbentuk melalui kluster berhierarki ini didasari dari analisis K-means sebelumnya, dimana jumlah kluster yang dikehendaki adalah tiga. Dalam hirarki terdapat beberapa ukuran kemiripan dan ketidakmiripan antar kluster yaitu: Pautan Tunggal, Pautan Lengkap, Pautan Centroid, Pautan Rataan, Pautan Median dan Ward. Pada metode ini digunakan pautan lengkap karena hasil yang didapatkan cukup baik karena tidak ada kelompok yang berdiri sendiri atau hanya terdiri satu provinsi pada kelompok.

TABEL 4. RATA-RATA VARIABEL BERDASARKAN PENGLASTERAN HIRARKI

Tingkat kerusakan lingkungan	IKU	IKA	IKTL	Sampah terkelola	Rata-rata
(1)	(2)	(3)	(4)	(5)	(6)
Tinggi	75,81	41,50	33,80	52,98	51,02
Sedang	88,77	55,45	55,96	44,91	61,28
Rendah	90,53	54,76	78,28	72,79	74,09

Pada kluster pertama hanya terdapat pada pulau jawa yang mana ada provinsi DKI Jakarta, Banten, Jawa barat dan DI Yogyakarta. Nilai rata-rata untuk semua indikator kerusakan lingkungan pada kluster pertama lebih rendah dibanding dengan nilai rata-rata provinsi di Indonesia dan juga paling rendah bila dibandingkan dengan rata-rata cluster lainnya. Dimana pada kelompok ini rata-rata gabungan dari IKU, IKA, IKTL, dan Sampah terkelola sebesar 51,02. Oleh karena itu, pada kluster pertama dapat dikatakan sebagai kluster yang memiliki tingkat kerusakan lingkungan yang paling tinggi dibanding cluster kedua dan ketiga.

Selanjutnya, pada pengelompokan kluster kedua terdiri dari provinsi yang cukup menyebar di berbagai pulau di Indonesia kecuali provinsi yang terletak di wilayah timur. Pada kelompok ini terdiri dari 17 provinsi. Daerah yang termasuk dalam kelompok ini memiliki nilai rata-rata gabungan IKU, IKA, IKTL

dan Sampah Terkelola yaitu 61,28. klaster kedua dapat dikatakan klaster yang memiliki karakteristik kerusakan lingkungan sedang karena nilai rata-rata semua indikator kerusakan lingkungan di cluster kedua masih relatif tidak terlalu jauh dari nilai rata-rata Indonesia. Pada kelompok ini karakteristiknya relatif berada di tengah-tengah jika dibandingkan pada kelompok tingkat kerusakan lingkungan tinggi dan rendah.

Untuk yang terakhir yaitu pada pengelompokan klaster ketiga, dimana terdiri dari provinsi yang secara geografis terletak di wilayah timur Indonesia ataupun berada di wilayah kalimantan. Pada klaster ini terdapat 13 provinsi yang termasuk. Daerah yang termasuk dalam kelompok ini memiliki nilai rata-rata gabungan IKU, IKA, IKTL dan Sampah Terkelola yaitu 74,09. Klaster ketiga dapat dikatakan klaster yang memiliki tingkat kerusakan lingkungan rendah karena rata-rata nilai karakteristik lebih tinggi dibandingkan klaster satu dan dan dua.

Metode Ensemble

Selanjutnya pengelompokan provinsi-provinsi di Indonesia berdasarkan kerusakan lingkungan dilakukan menggunakan metode ensemble. Banyaknya klaster yang terbentuk melalui klaster berhierarki ini didasari dari analisis K-means sebelumnya. Dimana pada jumlah klaster yang dikehendaki adalah tiga. Pada metode ini digunakan untuk mengkombinasikan sekumpulan solusi cluster dan dilakukan fungsi konsensus. Dalam hal ini fungsi konsensus yang digunakan adalah metode majority voting dimana setiap individu akan di voting berdasarkan metode K-means, Hirarki, DIANA, dan PAM. Sehingga berdasarkan voting tersebut individu tadi dapat dimasukan ke klaster dengan voting terbanyak.

TABEL 5. RATA-RATA VARIABEL BERDASARKAN PENKLASTERAN ENSEMBLE

Tingkat kerusakan lingkungan	IKU	IKA	IKTL	Sampah terkelola	Rata-rata
(1)	(2)	(3)	(4)	(5)	(6)
Tinggi	75,81	41,50	33,80	52,98	51,02
Sedang	88,87	55,35	55,53	46,37	61,53
Rendah	90,55	54,86	80,79	72,92	74,78

Pada klaster pertama hanya terdapat pada pulau jawa yang mana ada provinsi DKI Jakarta, Banten, Jawa barat dan DI Yogyakarta. Pada kelompok ini rata-rata gabungan dari IKU, IKA, IKTL, dan Sampah terkelola sebesar 51,02. Dapat dilihat juga rata-rata untuk semua indikator kerusakan lingkungan pada klaster pertama lebih rendah dibanding dengan nilai rata-rata provinsi di Indonesia dan juga paling rendah bila dibandingkan dengan rata-rata cluster lainnya. Oleh karena itu, pada klaster pertama dapat dikatakan sebagai klaster yang memiliki tingkat kerusakan lingkungan yang paling tinggi dibanding cluster kedua dan ketiga.

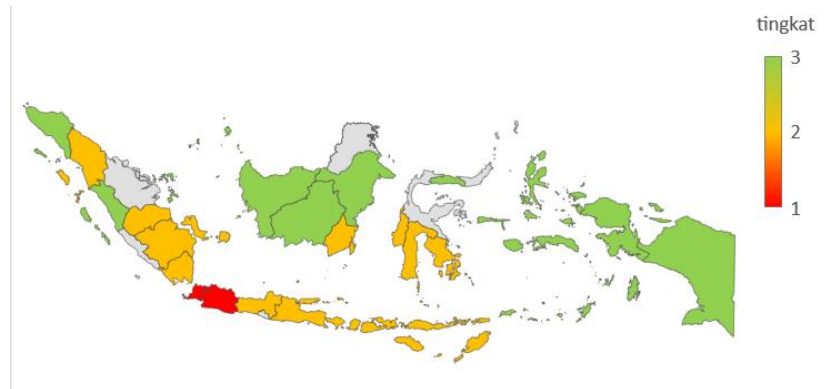
Selanjutnya, pada pengelompokan klaster kedua terdiri dari provinsi yang cukup menyebar di berbagai pulau di Indonesia kecuali provinsi yang terletak di wilayah timur. Pada kelompok ini terdiri dari 18 provinsi. Daerah yang termasuk dalam kelompok ini memiliki nilai rata-rata gabungan IKU, IKA, IKTL dan Sampah Terkelola yaitu 61,53. Klaster kedua dapat dikatakan klaster yang memiliki karakteristik kerusakan lingkungan sedang karena nilai rata-rata semua indikator kerusakan lingkungan di cluster kedua masih relatif tidak terlalu jauh dari nilai rata-rata Indonesia. Pada kelompok ini karakteristiknya relatif berada di tengah-tengah jika dibandingkan pada kelompok tingkat kerusakan lingkungan tinggi dan rendah.

Untuk yang terakhir yaitu pada pengelompokan klaster ketiga, dimana terdiri dari provinsi yang secara geografis terletak di wilayah timur Indonesia ataupun berada di wilayah kalimantan. Terdapat 12 provinsi yang termasuk klaster ini. Daerah yang termasuk dalam kelompok ini memiliki nilai rata-rata gabungan IKU, IKA, IKTL dan Sampah Terkelola yaitu 74,78. Klaster ketiga dapat dikatakan klaster yang memiliki tingkat kerusakan lingkungan rendah karena rata-rata nilai karakteristik lebih tinggi dibandingkan klaster satu dan dan dua.

Konsistensi pada Pengklasteran

Seperti yang telah dijelaskan sebelumnya bahwa metode pengklasteran *ensemble* didapatkan berdasarkan gabungan metode Hirarki pautan lengkap, K-means, DIANA, dan PAM. Dari setiap metode pengklasteran tersebut memiliki hasil yang berbeda-beda, oleh karena itu dibentuk klaster *ensemble* agar menciptakan satu hasil klaster yang terbaik. Pengklasteran dapat dikatakan konsisten apabila berdasarkan

keempat metode tersebut selaras dalam menggolongkan objek kedalam tingkat kerusakan lingkungan yang sama. Sehingga untuk melihat konsistensi pengklasteran pada pembentukan kluster ensemble maka berikut adalah hasilnya.



GAMBAR 5. PETA TEMATIK BERDASARKAN KONSISTENSI DARI METODE HIRARKI PAUTAN LENGKAP, K-MEANS, DIANA, DAN PAM

Pada kluster pertama dimana adalah tingkat kerusakan lingkungan tinggi hanya terdapat pada pulau jawa yang mana ada 3 provinsi yaitu: DKI Jakarta, Banten, dan Jawa Barat. Pada pengelompokan kluster kedua yang mana adalah daerah dengan kerusakan lingkungan sedang. Pada kelompok ini terdiri dari provinsi yang cukup menyebar dimana ada 14 provinsi yaitu : Bangka Belitung, Jambi, Lampung, Sumatera Selatan, Sumatera Utara, Jawa Tengah, Jawa Timur, Bali, Kalimantan Selatan, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Barat, Sulawesi Selatan, dan Sulawesi Tenggara. Selanjutnya kluster ketiga, dimana merupakan daerah yang tergolong tingkat kerusakan lingkungan rendah. Pada kelompok ini terdapat 12 provinsi yaitu: Aceh, Gorontalo, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Kepulauan Riau, Maluku, Maluku Utara, Papua, Papua Barat, dan Sumatera Barat.

Evaluasi Hasil Pengelompokan

Berdasarkan pengklasteran dengan beberapa metode yang telah dilakukan hingga akhirnya didapatkan 3 hasil pengklasteran. hasil kluster metode hirarki pautan lengkap, k-means dan metode cluster ensemble dibandingkan. Hasil pengelompokan dari metode hierarki, k-means, dan cluster ensemble dibandingkan menggunakan indeks Compactness (CP) dan indeks *Davies-Bouldin* (DB).

TABEL 6. HASIL EVALUASI DARI METODE PENGKLASTERAN

Metode	Davis Bouldin	Compactness
(1)	(2)	(3)
K-means	1.2042	1.9034
Hirarki	1.2082	1.9090
Ensemble	1.1779	1.9009

Hasil perbandingannya ditampilkan pada Tabel. Kriteria evaluasi kebaikan kluster, dimana metode yang mendapatkan nilai CP dan DB terkecil merupakan metode dengan hasil cluster yang lebih tepat. Pada tabel 4 terlihat bahwa metode *ensemble* memiliki nilai CP dan DB terkecil. Oleh karena itu, dapat disimpulkan bahwa dalam pengelompokan provinsi-provinsi di Indonesia berdasarkan tingkat kerusakan lingkungan di Indonesia, metode yang memberikan hasil lebih tepat adalah *cluster ensemble*.

IV. SIMPULAN DAN SARAN

Kesimpulan

1. Berdasarkan pada kasus ini, yaitu pengelompokan tingkat kerusakan lingkungan di Indonesia. Metode *ensemble* memberikan hasil pengklasteran yang terbaik dibandingkan metode *K-means* dan Hirarki berdasarkan evaluasi *compactness* dan *Davis bouldin*.
2. Berdasarkan metode pengklasteran *ensemble* didapatkan 4 provinsi termasuk ke dalam tingkat kerusakan lingkungan tinggi yaitu : DKI Jakarta, Banten, Jawa barat dan DI Yogyakarta. Untuk daerah yang termasuk tingkat kerusakan sedang ada 18 provinsi yaitu : Bali, Bangka Belitung, Bengkulu, Jambi, Jawa Tengah, Jawa Timur, Kalimantan Selatan, Lampung, Nusa Tenggara Barat, Nusa Tenggara Timur, Riau, Sulawesi Barat, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Utara, Sumatera Selatan, dan Sumatera Utara. Terdapat 12 provinsi dengan tingkat kerusakan lingkungan rendah yaitu : Aceh, Gorontalo, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Kepulauan Riau, Maluku, Maluku Utara, Papua, Papua Barat, dan Sumatera Barat.
3. Berdasarkan keempat metode pengklasteran dalam pembentukan *ensemble*, didapatkan ada 3 provinsi yang konsisten dengan tingkat kerusakan lingkungan tinggi; lalu ada 14 provinsi yang konsisten dengan tingkat kerusakan lingkungan sedang; dan terdapat 12 provinsi yang konsisten dengan tingkat kerusakan lingkungan rendah.

Saran

1. Bagi para masyarakat maupun pemangku kepentingan terutama pemerintah agar lebih memperhatikan lagi mengenai aspek lingkungan tidak hanya aspek ekonomi terutama pada daerah-daerah dengan tingkat kerusakan lingkungan tinggi.
2. Bagi peneliti selanjutnya dapat meneliti sebab akibat dari kerusakan lingkungan terutama pada daerah-daerah dengan kerusakan lingkungan tinggi.
3. Bagi peneliti selanjutnya agar dapat meneliti solusi dari masalah kerusakan lingkungan terutama pada daerah tingkat kerusakan lingkungan tinggi.

DAFTAR PUSTAKA

- [1] BPS. (2018). *Statistik Potensi Desa Indonesia 2018*. Badan Pusat Statistik.
- [2] Chiu, D. S., & Talhouk, A. (2018). diceR: an R package for class discovery using an ensemble driven approach. *BMC Bioinformatics*, 19(1), 11. <https://doi.org/10.1186/s12859-017-1996-y>
- [3] Croitoru, L., & Sarraf, M. (Eds.). (2010). *The Cost of Environmental Degradation*. The World Bank. <https://doi.org/10.1596/978-0-8213-8318-6>
- [4] Gujarati, D. N. (2003). Basic Econometrics. In *The Economic Journal*. Gary Burke. <https://doi.org/10.2307/2230043>
- [5] ILO. (2018). *The Employment Impact of Climate Change Adaptation. Input Document for the G20 Climate Sustainability Working Group* (p. 40). Retrieved from https://www.ilo.org/wcmsp5/groups/public/---ed_emp/documents/publication/wcms_645572.pdf
- [6] Johson RA, Winchern DW. 2007. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.
- [7] Landrigan, P. J., Fuller, R., Acosta, N. J. R., Adeyi, O., Arnold, R., Basu, N. (Nil), Baldé, A. B., Bertollini, R., Bose-O'Reilly, S., Boufford, J. I., Breyse, P. N., Chiles, T., Mahidol, C., Coll-Seck, A. M., Cropper, M. L., Fobil, J., Fuster, V., Greenstone, M., Haines, A., ... Zhong, M. (2018). The Lancet Commission on pollution and health. *The Lancet*, 391(10119), 462–512. [https://doi.org/10.1016/S0140-6736\(17\)32345-0](https://doi.org/10.1016/S0140-6736(17)32345-0)
- [8] Lelieveld, J., Klingmüller, K., Pozzer, A., Burnett, R. T., Haines, A., & Ramanathan, V. (2019). Effects of fossil fuel and total anthropogenic emission removal on public health and climate. *Proceedings of the National Academy of Sciences*, 116(15), 7192–7197. <https://doi.org/10.1073/pnas.1819989116>
- [9] Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). Understanding of Internal Clustering Validation Measures. *2010 IEEE International Conference on Data Mining*, 911–916. <https://doi.org/10.1109/ICDM.2010.35>
- [10] Mattjik AA, Sumertajaya IM. 2011. *Sidik Peubah Ganda dengan Menggunakan SAS*. Wibawa GNA, Hadi AF, editor. Bogor (ID): IPB Press
- [11] Nicolas Maitre, Guillermo Montt and Catherine Saget, with additional contributions by Christoph Ernst, María Teresa Gutiérrez, Takaaki Kizu, Tahmina Karimova, M., & Lieuw-Kie-Song, and M. T. (n.d.). *The employment impact of climate change adaptation Input Document for the G20 Climate Sustainability Working Group*.
- [12] Noor N. M., Yahaya A. S., Ramli N. A., Luca F. A., Abdullah M. M. A., Sandu A. V. (2015), Variation of Air Pollutant (Particulate Matter - PM10) in Malaysia. Study in the Southwest Coast of Peninsular Malaysia, *Revista de Chimie* 66(9): 1443-1447.

- [13] Solarin, S. A. (2019). Convergence in CO₂ emissions, carbon footprint and ecological footprint: evidence from OECD countries. *Environmental Science and Pollution Research*, 26(6), 6167–6181. <https://doi.org/10.1007/s11356-018-3993-8>
- [14] Strehl A, Gosh J. 2012. A Knowledge Reuse Framework for Combining Partitionings. *The Journal of Machine Learning Research*. 3(1):583-6
- [15] World Bank Group, & IHME. (2016). *The cost of air pollution: Strengthening the Economic Case for Action. The World Bank and Institute for Health Metrics and Evaluation University of Washington, Seattle* (pp. 1–102). Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/000368497326688#.Va9xXPnQjVI>