



Aplikasi Metode Ward dengan Berbagai Pengukuran Jarak (Studi Kasus: Klasifikasi Tingkat Perekonomian di Indonesia)

Rhavida Anniza Andyani^{1✉}, Muhammad Qolbi Shobri², Muhamad Adzib Baihaqi³,
Putri Balqis Al-Kubro⁴, Muhammad Syahriandi Adhantoro⁵

¹⁻⁴Fakultas Ilmu Formal Ilmu Terapan, Universitas Muhammadiyah Madiun, Indonesia

⁵Fakultas Komunikasi dan Informatika, Universitas Muhammadiyah Surakarta, Indonesia

✉Korespondensi Penulis

Rhavida Anniza Andyani

Fakultas Ilmu Formal Ilmu Terapan, Universitas Muhammadiyah Madiun, Indonesia

raa250@ummad.ac.id

doi: 10.56972/jikm.v4i2.208

Submit: 15 Juni 2024 | Revisi: 14 Oktober 2024 | Diterima: 17 Oktober 2024

Dipublikasikan: 28 Oktober 2024 | Periode Terbit: Oktober 2024

Abstrak

Metode Ward merupakan metode *hierarchical clustering* di mana jarak digunakan sebagai acuan dalam proses klusterisasi. Penelitian ini menggunakan 5 jarak yang berbeda yang diaplikasikan pada tujuh variabel yang berkaitan tentang ekonomi, yakni Produk Domestik Regional Bruto (PDRB), Pengeluaran Pemerintah, Penanaman Modal Dalam Negeri (PMDN), Penanaman Modal Asing (PMA), Tingkat Partisipasi Angkatan Kerja (TPAK), Indeks Pembangunan Manusia (IPM), dan Rasio Elektrifikasi. Data yang digunakan diperoleh dari data BPS dari 34 provinsi tahun 2022. Tujuannya mengetahui jarak optimal dalam penerapan klasifikasi dengan menggunakan metode Ward pada data ekonomi. Selain itu untuk mengetahui provinsi mana saja yang tergolong dalam tingkat ekonomi rendah, sedang, dan tinggi. Dengan menggunakan nilai *Agglomerative Coefficients* menunjukkan bahwa jarak squared Euclidean bernilai paling tinggi yang dapat diartikan merupakan jarak paling optimal diantara kelima jarak yang digunakan (*Euclidean*, *Squared Euclidean*, *Manhattan*, *Minkowski*, dan *Canberra*). Tiga klaster yang dihasilkan dengan menggunakan jarak squared *Euclidean* adalah klaster 1 merupakan provinsi dengan tingkat ekonomi rendah yang terdiri dari 24 provinsi, klaster 2 terdiri dari 9 provinsi tergolong dalam tingkat ekonomi menengah, dan klaster 3 terdiri dari 1 provinsi, yakni DKI Jakarta yang termasuk dalam perekonomian tinggi.

Kata Kunci: *agglomerative coefficients*, indeks pembangunan manusia, penanaman modal dalam negeri, penanaman modal asing, produk domestik regional bruto, squared euclidean, tingkat partisipasi angkatan kerja

1. Pendahuluan

Analisis kluster adalah metode statistik yang digunakan untuk mengelompokkan objek-objek berdasarkan kesamaan karakteristik tertentu. Dalam analisis kluster, objek yang memiliki kesamaan yang lebih besar akan dikelompokkan dalam satu kelompok, sementara objek yang memiliki perbedaan yang lebih besar akan dimasukkan ke dalam kelompok yang berbeda. Tujuan utama dari analisis kluster adalah untuk mengidentifikasi struktur atau pola dalam data yang tidak diketahui sebelumnya (Johnson & Wichern, 2007). Konsep dasar analisis kluster adalah pengelompokan objek-objek yang memiliki kesamaan tertentu, sementara objek-objek dalam kelompok yang berbeda akan menunjukkan perbedaan yang signifikan (Johnson & Wichern, 2007). Hal ini membuat analisis kluster sangat berguna dalam banyak aplikasi, seperti segmentasi pasar, pengelompokan pelanggan, analisis sosial, dan pengelompokan geografis.

Secara umum, analisis kluster dapat diterapkan pada data yang sangat besar dengan banyak variabel. Data tersebut dapat memiliki berbagai skala, seperti ordinal, interval, maupun rasio, yang memungkinkan analisis kluster untuk diterapkan pada berbagai jenis data (Talakua, Leleury, & Talluta, 2017). Salah satu keunggulan utama dari analisis kluster adalah kemampuannya untuk mengelompokkan data tanpa memerlukan label atau kategori sebelumnya, yang menjadikannya sangat fleksibel dalam eksplorasi

data (Pradana et al., 2024). Dalam praktiknya, terdapat dua jenis utama dalam analisis kluster, yaitu metode hierarki dan metode non-hierarki. Metode hierarki memberikan kemudahan dalam visualisasi hasil pengelompokan melalui dendrogram, yang menggambarkan bagaimana objek-objek dikelompokkan secara bertahap (Setiawan et al., 2020). Metode non-hierarki, di sisi lain, umumnya lebih efisien untuk data yang lebih besar dan kompleks, seperti metode K-Means yang sering digunakan dalam banyak aplikasi.

Terdapat berbagai metode yang digunakan dalam analisis kluster, dan masing-masing metode memiliki kelebihan dan kekurangannya. Dalam klusterisasi hierarki, ada beberapa teknik yang sering digunakan, seperti Single Linkage, Complete Linkage, Average Linkage, dan Ward (Setiawan et al., 2020). Metode Ward adalah salah satu teknik klusterisasi yang populer, yang berfokus pada minimisasi varian dalam kelompok dan meminimalkan kesalahan kuadrat jumlah (Sum of Squares Error / SSE). Keunggulan utama metode Ward adalah kemampuannya untuk menghasilkan pembagian kluster yang lebih homogen dan lebih terstruktur, yang menjadikannya efektif dalam banyak aplikasi analisis data (Gundono, 2011).

Banyak penelitian yang telah membandingkan berbagai metode dalam analisis kluster, terutama dalam konteks aplikasi dunia nyata. Sebagai contoh, Irwan et al. (2024) melakukan perbandingan antara metode Complete Linkage dan Ward

untuk mengklasifikasikan Indeks Pembangunan Manusia (IPM) di Kota/Kabupaten Provinsi Sulawesi tahun 2022 dan menemukan bahwa metode Ward memberikan hasil yang lebih baik dalam hal akurasi klasifikasi. Begitu juga Imasdiani et al. (2022) yang membandingkan beberapa metode klaster, dan hasilnya menunjukkan bahwa metode Ward memberikan nilai simpangan baku yang lebih kecil baik dalam maupun antar kelompok, yang menunjukkan kualitas klaster yang lebih baik dibandingkan dengan Average Linkage. Penelitian lain yang dilakukan oleh Pradana et al. (2023) dan Prayitno et al. (2024) juga menunjukkan bahwa metode Ward memiliki kinerja terbaik dalam mengklasifikasikan data kinerja kantor unit bank ABC dibandingkan dengan metode Nearest Neighbor dan K-Means. Temuan-temuan ini menunjukkan bahwa metode Ward memiliki keunggulan dalam hal kualitas pengelompokan yang lebih stabil dan lebih akurat, terutama pada data yang lebih kompleks dan multivariat.

Dasar dari pengklasifikasian dalam analisis klaster adalah ukuran kemiripan antar objek yang diukur melalui jarak antar objek tersebut (Utami et al., 2024). Ukuran jarak yang digunakan dalam analisis klaster dapat beragam, seperti Euclidean, Squared Euclidean, Minkowski, Manhattan, dan Canberra. Ukuran jarak ini bertujuan untuk mengukur seberapa dekat atau jauh dua objek berdasarkan variabel-variabel yang ada. Ukuran Euclidean merupakan ukuran jarak yang paling sering digunakan karena sifatnya yang sederhana dan mudah dipahami. Namun, beberapa penelitian juga membandingkan

keefektifan berbagai ukuran jarak dalam menghasilkan klaster yang baik. Misalnya, Nishom (2019) melakukan perbandingan antara ukuran jarak Euclidean, Minkowski, dan Manhattan pada algoritma K-Means, dan menemukan bahwa ukuran jarak Euclidean memberikan hasil terbaik. Hal serupa juga ditemukan dalam penelitian oleh Singh et al. (2013), yang menunjukkan bahwa jarak Euclidean lebih unggul dibandingkan dengan Manhattan dan Minkowski dalam algoritma K-Means. Meskipun demikian, beberapa penelitian juga menunjukkan bahwa pemilihan ukuran jarak yang tepat sangat bergantung pada karakteristik data yang dianalisis.

Namun, perbandingan ukuran jarak dalam analisis klaster hierarki, terutama pada metode Ward, masih terbatas. Oleh karena itu, penelitian ini akan mengeksplorasi penerapan metode Ward dengan menggunakan berbagai ukuran jarak untuk melihat mana yang memberikan hasil klasifikasi yang lebih optimal. Kelima ukuran jarak yang digunakan dalam penelitian ini adalah Euclidean, Squared Euclidean, Manhattan, Minkowski, dan Canberra. Penelitian ini bertujuan untuk memperbandingkan efektivitas masing-masing ukuran jarak dalam menghasilkan struktur klaster yang baik, dengan menggunakan data perekonomian Indonesia sebagai objek analisis. Ketimpangan ekonomi antar provinsi di Indonesia menjadi fokus utama, karena meskipun ada berbagai provinsi dengan pertumbuhan ekonomi yang tinggi, banyak daerah yang masih tertinggal secara signifikan. Pengklasifikasian ini bertujuan untuk

mengidentifikasi provinsi-provinsi dengan tingkat perekonomian yang rendah, menengah, dan tinggi, serta untuk memberikan wawasan bagi pemerintah dalam merumuskan kebijakan yang lebih tepat untuk mengurangi ketimpangan ekonomi dan meningkatkan kesejahteraan masyarakat di seluruh Indonesia (Suleiman et al., 2019).

Penelitian ini menggunakan tujuh variabel ekonomi utama sebagai dasar klasifikasi, yaitu Produk Domestik Regional Bruto (PDRB), Pengeluaran Pemerintah, Penanaman Modal Dalam Negeri (PMDN), Penanaman Modal Asing (PMA), Tingkat Partisipasi Angkatan Kerja (TPAK), Indeks Pembangunan Manusia (IPM), dan Rasio Elektrifikasi. Variabel-variabel ini dipilih karena mereka mencerminkan faktor-faktor yang mempengaruhi perkembangan ekonomi suatu provinsi. PDRB, misalnya, mencerminkan hasil kegiatan ekonomi suatu daerah, sementara TPAK dan IPM mewakili kualitas sumber daya manusia yang ada. Pengeluaran pemerintah, PMDN, dan PMA juga merupakan indikator penting dalam menilai potensi ekonomi dan investasi di suatu wilayah (Andyani, Setiawan, & Ratnasari, 2023). Data yang digunakan dalam penelitian ini berasal dari Badan Pusat Statistik (BPS) yang mencakup 34 provinsi di Indonesia pada tahun 2022.

Penelitian ini diharapkan dapat memberikan kontribusi penting dalam pemahaman mengenai ketimpangan ekonomi di Indonesia dan memberikan rekomendasi kebijakan yang lebih ber-

basis data dalam upaya pemerataan pembangunan ekonomi antar provinsi. Dengan menggunakan metode Ward dan berbagai ukuran jarak, penelitian ini akan memperdalam pemahaman mengenai struktur perekonomian Indonesia dan mengidentifikasi faktor-faktor utama yang perlu diperhatikan dalam merumuskan kebijakan pembangunan yang lebih inklusif dan berkelanjutan.

2. Metode

a. Asumsi Analisis Klaster

Pengecekan data mengenai asumsi yang harus terpenuhi sebelum data diolah dengan menggunakan analisis klasifikasi adalah sampel yang digunakan representatif dan tidak terjadi kasus multikolinieritas. Pengujian untuk menentukan bahwa sampel yang digunakan sudah representatif dalam mewakili populasi yang ada menggunakan Uji *Kaiser Meyer Olkin* (KMO).

KMO merupakan indeks yang menyatakan ukuran ketepatan sampel pada setiap indikator telah tercukupi. Sampel dikatakan representatif apabila KMO menunjukkan nilai antara 0.5 hingga 1 (Apriliana & Widodo, 2023).

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (1)$$

keterangan:

p : banyaknya variabel

r_{ij} : koefisien korelasi variabel ke- i dan j

a_{ij} : koefisien korelasi parsial variabel ke- i dan j , (Irwani, Sanusi, & Hasanah, 2024).

Tidak terjadi kasus multikolinieritas merupakan asumsi kedua yang dalam analisis kluster. Multikolinieritas adalah kondisi di mana variabel prediktor terkait erat satu sama lain. Adanya kasus multikolinieritas menyebabkan ketidaktepatan pada estimator (Greene, 2012). Multikolinieritas menyebabkan kesulitan dalam melihat pengaruh variabel prediktor terhadap variabel respon. Multikolinieritas terjadi jika estimasi menunjukkan nilai R^2 yang tinggi (lebih dari 0,8), nilai F yang tinggi, dan nilai t statistik tidak signifikan untuk semua atau sebagian besar variabel (Gujarati, 2004).

Variance Inflation Factor (VIF) merupakan indikator untuk mendeteksi adanya kasus multikolinieritas. Jika VIF bernilai lebih dari 10 maka dikatakan terdapat kasus multikolinieritas. Semakin tinggi nilai VIF, maka semakin tinggi pula korelasi antar variabel (Apriliana & Widodo, 2023).

$$VIF = \frac{1}{1 - R_k^2} \quad (2)$$

R_k^2 : koefisien korelasi antara 1 variabel prediktor dengan variabel prediktor lainnya, dimana $k=1,2,3,\dots,p$.

b. Metode Ward

Metode Ward didasarkan pada kriteria *Sum Square Error* (SSE), adalah teknik pengelompokan hierarki aglomeratif yang menggunakan variansi sekecil mungkin untuk menghasilkan kluster. Metode ini meminimalkan variansi dalam kelompok, dan biasanya digunakan untuk menggabungkan kelompok dengan

jumlah kecil. (Apriliana & Widodo, 2023). Pembentukan jarak SSE dapat dihitung apabila kluster memiliki elemen lebih dari satu objek, jika hanya memiliki satu objek, pembentukan jarak SSE adalah nol (Gundono, 2011). Formula SSE dapat dituliskan pada persamaan 3.

$$SSE = \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})'(\mathbf{x}_i - \bar{\mathbf{x}}) \quad (3)$$

c. Standarisasi Data

Standarisasi data merupakan proses transformasi nilai variabel asli menjadi nilai baru dengan mean 0 dan standar deviasi 1 (Supranto, 2010). Standarisasi data dilakukan ketika data yang digunakan memiliki skala atau satuan yang berbeda. Standarisasi di formulakan dengan persamaan 4.

$$Z = \frac{x - \bar{x}}{s} \quad (4)$$

$\bar{x}$: rata-rata data

x : nilai data

s : simpangan baku

d. Ukuran Jarak

Jarak digunakan sebagai ukuran dalam pengklasifikasian. Terdapat berbagai metode pengukuran jarak yang dapat digunakan dalam analisis kluster, seperti:

1) Jarak Euclidean

Jarak *Euclidean* adalah jarak yang paling sering digunakan dalam analisis kluster. Jarak *Euclidean* antara kluster ke- i dan ke- j dari p peubah dirumuskan pada persamaan 5.

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

2) Jarak Squared Euclidean

Squared Euclidean adalah jumlah kuadrat perbedaan deviasi nilai masing-masing variabel (Hair, Anderson, Thatham, & Black, 2010). Jarak ini merupakan variasi dari jarak Euclidean. Persamaan 6 mendefinisikan jarak *Squared Euclidean* antara kelompok objek ke-*i* dan kelompok objek ke-*j* dari variabel *p*.

$$d_{ij} = \sum_{k=1}^p (x_{ik} - x_{jk})^2 \quad (6)$$

3) Jarak Manhattan atau City Block

Jarak *Manhattan* dihitung sebagai jumlah nilai perbedaan mutlak dari setiap variabel. Dalam beberapa situasi, jarak dari *Manhattan* menghasilkan jarak yang sebanding dengan jarak *Euclidean*. Kelebihan dari jarak ini adalah dapat mendeteksi *outlier* dengan baik. Jarak *Manhattan* dapat dirumuskan dalam persamaan 7.

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}| \quad (7)$$

4) Jarak Minkowski

Minkowski adalah generalisasi dari jarak *Euclidean* dan *Manhattan*, dan merupakan metrik dalam ruang vektor di mana suatu norma didefinisikan (*normed vector space*) (Nishom, 2019).

$$d_{ij} = \sum_{i=1}^n \left(|x_{ik} - x_{jk}|^p \right)^{1/p} \quad (8)$$

5) Jarak Canberra

Metode pengukuran jarak *Canberra* digunakan untuk menghitung jarak antara dua titik, di mana data yang digunakan adalah data asli dan berada dalam ruang vector .

$$d_{ij} = \sum_{k=1}^p \left(\frac{|x_{ik} - x_{jk}|}{|x_{ik} + x_{jk}|} \right) \quad (9)$$

e. Agglomerative Coefficients

Agglomerative coefficients (AC) menunjukkan nilai optimum dalam menghasilkan pengelompokan yang ideal. Nilai ini digunakan untuk menilai kekuatan struktur pengelompokan. Nilai 0–1 dari koefisien menunjukkan bahwa struktur pengelompokan lebih seimbang dan baik, sedangkan nilai 0 menunjukkan bahwa algoritma belum menemukan struktur alami atau *cluster* yang terbentuk dengan buruk (Vandevor, 2020).

3. Hasil dan Pembahasan

Sebelum melakukan analisis kluster, langkah pertama yang perlu dilakukan adalah memastikan bahwa data yang digunakan memenuhi asumsi dasar yang diperlukan dalam analisis kluster, yaitu sampel yang representatif dan tidak terjadi multikolinieritas antar variabel. Kedua asumsi ini sangat penting untuk memastikan bahwa hasil yang diperoleh dapat dipertanggungjawabkan secara statistik dan tidak mengarah pada kesimpulan yang bias atau tidak valid.

a. Pengujian Asumsi Represen-tativi-tas Sampel

Dalam penelitian ini, pengujian representativitas sampel dilakukan dengan menggunakan Uji Kaiser-Meyer-Olkin (KMO). Uji KMO bertujuan untuk mengukur seberapa baik variabel-variabel dalam dataset saling berkorelasi dan dapat mewakili populasi yang lebih besar. Nilai KMO yang lebih tinggi menunjukkan bahwa data lebih cocok untuk analisis faktor atau kluster.

Hasil pengujian dengan menggunakan uji KMO menunjukkan nilai sebesar **0.722**, yang berada di atas nilai ambang batas minimum yaitu 0.5. Hal ini menunjukkan bahwa sampel yang digunakan dalam penelitian ini dapat mewakili populasi dengan baik, dan data dapat dilanjutkan untuk dianalisis lebih lanjut menggunakan teknik analisis kluster. Nilai KMO yang berada di kisaran 0.7-0.8 umumnya menunjukkan tingkat kecocokan yang baik antara variabel-variabel yang ada dalam data, sehingga model klusterisasi yang akan diterapkan dapat diandalkan.

b. Pengujian Multikolinieritas

Langkah selanjutnya adalah melakukan uji multikolinieritas untuk memastikan tidak ada hubungan yang terlalu kuat antara variabel-variabel independen yang dapat mempengaruhi kestabilan dan ketepatan estimasi dalam model. Multikolinieritas dapat menyebabkan masalah dalam interpretasi pengaruh masing-masing variabel, karena variabel-variabel yang

sangat berkorelasi dapat saling tumpang tindih dalam menjelaskan variabilitas data.

Untuk mendeteksi adanya multikolinieritas, digunakan Variance Inflation Factor (VIF). VIF mengukur seberapa besar inflasi (kenaikan) varians estimasi koefisien regresi ketika variabel independen berkorelasi dengan variabel independen lainnya. Secara umum, nilai VIF yang lebih besar dari 10 menunjukkan adanya multikolinieritas yang signifikan. Nilai VIF pada masing-masing variabel dapat dilihat Tabel 1.

Tabel 1. Koefisien *Variance Inflation Factor* (VIF)

X1	X2	X3	X4	X5	X6	X7
2.056	3.215	3.685	1.621	1.225	1.600	1.161

Hasil perhitungan VIF untuk ketujuh variabel yang digunakan dalam analisis ini menunjukkan bahwa semua nilai VIF berada di bawah angka 10, yang berarti bahwa tidak ada variabel yang menunjukkan adanya multikolinieritas yang signifikan. Oleh karena itu, ketujuh variabel yang digunakan dalam penelitian ini dapat diterima untuk dilanjutkan dalam analisis kluster tanpa risiko multikolinieritas yang dapat mengganggu hasil analisis.

c. Analisis Kluster dengan Metode Ward

Setelah memastikan bahwa data memenuhi asumsi dasar, langkah berikutnya adalah melakukan analisis kluster untuk

mengelompokkan provinsi-provinsi di Indonesia berdasarkan variabel-variabel yang dipilih. Variabel-variabel yang digunakan dalam analisis kluster ini meliputi Produk Domestik Regional Bruto (PDRB), Pengeluaran Pemerintah, Penanaman Modal Dalam Negeri (PMDN), Penanaman Modal Asing (PMA), Tingkat Partisipasi Angkatan Kerja (TPAK), Indeks Pembangunan Manusia (IPM), dan Rasio Elektrifikasi. Ketujuh variabel ini dianggap memiliki keterkaitan erat dalam menggambarkan kondisi ekonomi suatu provinsi dan digunakan untuk mengklasifikasikan provinsi-provinsi tersebut menjadi tiga kelompok berdasarkan tingkat perekonomiannya, yaitu rendah, menengah, dan tinggi.

Dalam analisis ini, digunakan metode Ward yang merupakan salah satu teknik pengelompokan hierarkis aglomeratif. Metode Ward memiliki keunggulan dalam meminimalkan varians dalam kluster, sehingga hasil klasterisasi yang diperoleh cenderung lebih homogen dan representatif (Ayu et al., 2021). Dalam implementasinya, metode Ward menggunakan ukuran jarak untuk mengukur kedekatan antar provinsi, yang pada gilirannya akan membentuk kluster-kluster berdasarkan kemiripan antar provinsi (Yamin et al., 2020).

Sebagai bagian dari evaluasi kualitas pengelompokan, digunakan lima ukuran jarak yang berbeda, yaitu Euclidean, Squared Euclidean, Manhattan, Minkow-

ski, dan Canberra. Tujuannya adalah untuk mengidentifikasi ukuran jarak mana yang memberikan hasil klasterisasi yang optimal (Hafida et al., 2020). Untuk mengevaluasi kekuatan hasil pengelompokan, digunakan Agglomerative Coefficients (AC), yang menunjukkan seberapa seimbang dan baik struktur kluster yang terbentuk. Nilai AC yang mendekati 1 menunjukkan bahwa kluster yang terbentuk memiliki struktur yang baik dan seimbang, sedangkan nilai AC yang rendah menunjukkan bahwa pengelompokan yang terbentuk kurang optimal.

d. Hasil Pengklasifikasian dengan Metode Ward

Tabel 2 menunjukkan nilai Agglomerative Coefficients (AC) untuk setiap ukuran jarak yang digunakan dalam analisis kluster. Berdasarkan hasil yang diperoleh, ukuran jarak Squared Euclidean memberikan nilai AC tertinggi di antara kelima ukuran jarak yang diterapkan (Fauziyah et al., 2022). Hal ini mengindikasikan bahwa pengelompokan yang dilakukan dengan menggunakan ukuran jarak Squared Euclidean memberikan struktur kluster yang lebih baik dan lebih seimbang dibandingkan dengan ukuran jarak lainnya (Odewumi et al., 2019). Nilai AC yang tinggi ini menunjukkan bahwa data yang dikelompokkan dengan metode Ward menggunakan ukuran jarak Squared Euclidean sangat cocok untuk membentuk kluster-kluster yang jelas dan bermakna (Ansari et al., 2022).

Gambar 1 menunjukkan dendrogram hasil klasterisasi dengan menggunakan metode Ward dan ukuran jarak Squared Euclidean. Dendrogram ini memperlihatkan bagaimana provinsi-provinsi dikelompokkan secara hierarkis berdasarkan kesamaan karakteristik ekonomi mereka. Hasil klasterisasi ini menghasilkan tiga klaster utama, yang masing-masing menggambarkan tingkat perekonomian yang berbeda, yaitu klaster dengan tingkat perekonomian rendah, menengah, dan tinggi.

Berdasarkan dendrogram tersebut, hasil klasterisasi menunjukkan bahwa Klaster 1 terdiri dari 24 provinsi, yang semuanya memiliki karakteristik ekonomi yang lebih rendah. Klaster 2 terdiri dari 9 provinsi yang memiliki perekonomian menengah, sementara Klaster 3 terdiri dari hanya satu provinsi, yaitu DKI Jakarta, yang memiliki tingkat perekonomian tertinggi di antara semua provinsi di Indonesia.

Nilai AC pada kelima ukuran jarak yang diterapkan dalam metode Ward dapat disajikan pada Tabel 2.

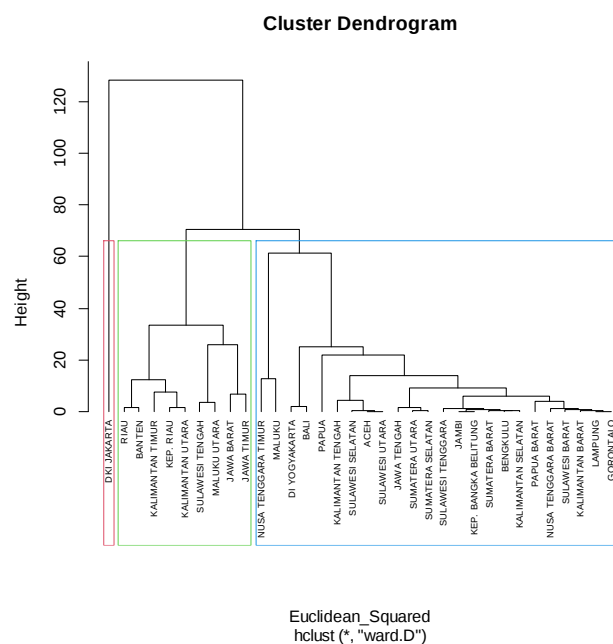
Tabel 2. Nilai Agglomerative Coefficients (AC)

Ukuran Jarak	AC
<i>Euclidean</i>	0.878
<i>Squared Euclidean</i>	0.947
<i>Manhattan</i>	0.928
<i>Minkowski</i>	0.878
<i>Canberra</i>	0.899

Dari kelima metode pengukuran jarak, dapat diketahui bahwa *Euclidean squared* menunjukkan nilai AC tertinggi.

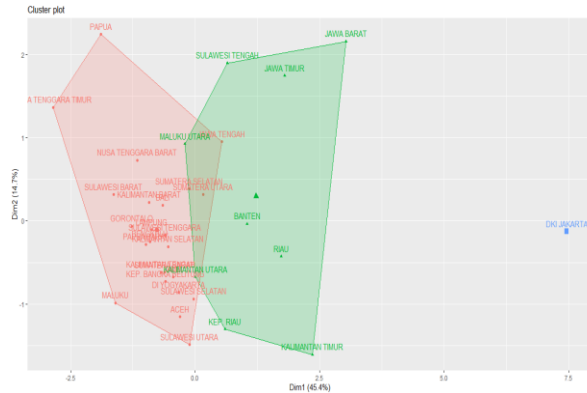
Artinya bahwa dengan menggunakan metode Ward dengan menggunakan jarak *Euclidean squared* membentuk struktur pengelompokan yang sangat baik, hal ini ditunjukkan dengan nilai AC yang mendekati 1.

Dendrogram hasil klasifikasi menggunakan metode Ward dengan ukuran jarak *Squared Euclidean* dapat dilihat pada Gambar 1.



Gambar 1. Dendrogram Klaster dengan Metode Ward dan Jarak Euclidean Squared

Gambar 1 menunjukkan hasil klasterisasi dengan 3 kelompok, dengan hasil pada klaster 1 berjumlah 24 provinsi, klaster 2 berjumlah 9 provinsi, klaster 3 berjumlah 1 provinsi. Gambar 2 menunjukkan hasil distribusi provinsi untuk setiap kelompok.



Gambar 2. Clustering Plot

Berdasarkan hasil pada Gambar 2, provinsi yang masuk dalam ketiga kelompok dapat dijabarkan sebagai berikut:

- Aceh, Sumatera Utara, Sumatera Barat, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Jawa Tengah, DI Yogyakarta, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Papua Barat, dan Papua;
- Riau, Kep. Riau, Jawa Barat, Jawa Timur, Banten, Kalimantan Timur, Kalimantan Utara, Sulawesi Tengah, Maluku Utara;
- DKI Jakarta.

Tabel 3. Rata-Rata Nilai Variabel pada Setiap Klaster

Variabel	Klaster 1	Klaster 2	Klaster 3
PDRB	32039.440	66202.156	183598.470
Pengeluaran Pemerintah	17161175.648	25253462.271	233899196.870
PMDN	7663.438	31069.222	89223.600
PMA	476.192	3381.344	3744.100
TPAK	70.090	67.796	65.210
IPM	71.128	73.132	81.650
Rasio Elektrifikasi	99.175	99.988	99.990

e. Distribusi Provinsi dalam Setiap Klaster

Hasil distribusi provinsi untuk masing-masing klaster dapat dilihat pada Gambar 2. Pembagian ini memberikan gambaran lebih rinci mengenai penyebaran provinsi berdasarkan tingkat perekonomian mereka. Secara umum, hasil klasifikasi ini menunjukkan ketimpangan ekonomi yang cukup signifikan antar provinsi di Indonesia.

1) Klaster 1 (Ekonomi Rendah)

Terdapat 24 provinsi yang masuk dalam klaster ini, yang sebagian besar merupakan provinsi-provinsi di wilayah luar Jawa. Provinsi-provinsi ini umumnya memiliki nilai rata-rata terendah untuk hampir semua variabel ekonomi yang digunakan dalam analisis, kecuali untuk Tingkat Partisipasi Angkatan Kerja (TPAK), yang menunjukkan angka yang relatif lebih tinggi. Hal ini mengindikasikan bahwa meskipun sebagian besar penduduk di klaster ini aktif dalam

kegiatan ekonomi, namun secara keseluruhan, perekonomian mereka tetap berada pada tingkat rendah.

2) **Klaster 2 (Ekonomi Menengah)**

Terdapat 9 provinsi yang masuk dalam klaster ini, yang memiliki karakteristik perekonomian menengah. Provinsi-provinsi ini menunjukkan nilai rata-rata yang lebih tinggi dibandingkan dengan Klaster 1, tetapi masih jauh di bawah Klaster 3. Peningkatan pengeluaran pemerintah di klaster ini juga menjadi indikator bahwa provinsi-provinsi dalam klaster ini tengah mengalami pengembangan infrastruktur dan layanan publik yang signifikan untuk mendukung pertumbuhan ekonomi.

3) **Klaster 3 (Ekonomi Tinggi)**

Hanya DKI Jakarta yang masuk dalam klaster ini. Provinsi ini menunjukkan nilai tertinggi untuk hampir semua variabel ekonomi yang digunakan dalam analisis, kecuali **Pengeluaran Pemerintah** dan **TPAK**. Hal ini menunjukkan bahwa meskipun Jakarta memiliki perekonomian yang sangat kuat, persentase penduduk usia kerja yang terlibat aktif dalam ekonomi relatif lebih rendah dibandingkan dengan provinsi lainnya. Ini bisa jadi disebabkan oleh tingginya tingkat urbanisasi dan ketergantungan pada sektor-sektor yang membutuhkan lebih sedikit tenaga kerja langsung.

f. **Karakteristik Setiap Klaster**

Berdasarkan rata-rata nilai setiap variabel dalam masing-masing klaster, dapat disimpulkan bahwa:

1) **Klaster 1 (Ekonomi Rendah)**

Karakteristik utama dari klaster ini adalah tingkat pengeluaran pemerintah yang rendah, rendahnya IPM dan PDRB, serta rendahnya jumlah PMA dan PMDN. Meskipun TPAK tinggi, ini mencerminkan tingginya jumlah penduduk yang aktif dalam sektor-sektor ekonomi yang kurang produktif atau bergaji rendah.

2) **Klaster 2 (Ekonomi Menengah)**

Klaster ini menunjukkan nilai variabel yang berada di antara klaster 1 dan 3, dengan PDRB, PMA, dan PMDN yang lebih tinggi dibandingkan klaster 1, namun belum mencapai tingkat yang ditemukan di klaster 3. Provinsi dalam klaster ini umumnya sedang dalam tahap transisi ekonomi, dengan peningkatan signifikan pada infrastruktur dan investasi.

3) **Klaster 3 (Ekonomi Tinggi)**

Jakarta sebagai satu-satunya provinsi dalam klaster ini menunjukkan bahwa pusat ekonomi Indonesia terkonsentrasi di wilayah ini, dengan pengeluaran pemerintah yang tinggi namun lebih sedikit partisipasi aktif dari angkatan kerja.

4. **Simpulan**

Hasil analisis klaster dengan metode Ward menunjukkan adanya ketimpangan perekonomian antar provinsi di Indonesia,

yang terbagi dalam tiga klaster: ekonomi rendah, ekonomi menengah, dan ekonomi tinggi. Klaster ekonomi rendah terdiri dari 24 provinsi dengan perekonomian yang relatif lemah, meskipun tingkat partisipasi angkatan kerja tinggi. Klaster ekonomi menengah berisi 9 provinsi yang menunjukkan perkembangan ekonomi yang lebih baik, namun masih berada di bawah klaster ekonomi tinggi yang hanya mencakup DKI Jakarta. DKI Jakarta sebagai pusat ekonomi Indonesia menunjukkan nilai tertinggi dalam berbagai variabel ekonomi, meskipun TPAK-nya lebih rendah. Temuan ini mengindikasikan perlunya kebijakan pemerataan pembangunan untuk mengurangi ketimpangan ekonomi antar provinsi, terutama dengan meningkatkan infrastruktur, investasi, dan kualitas sumber daya manusia di provinsi dengan ekonomi rendah dan menengah. Rekomendasi untuk penelitian masa depan adalah untuk mengeksplorasi faktor-faktor lain yang memengaruhi ketimpangan ekonomi, seperti kebijakan fiskal, distribusi sumber daya alam, serta faktor sosial dan politik yang dapat berkontribusi pada peningkatan atau penurunan kesejahteraan antar provinsi, serta untuk menguji penggunaan metode klasterisasi lain atau variabel tambahan dalam rangka memperbaiki akurasi klasifikasi ekonomi antar wilayah.

5. Daftar Pustaka

- Andyani, R. A., Setiawan, & Ratnasari, V. (2023). Estimation of Random Effect Probit Panel Parameter Using Adaptive Gauss Hermite Quadrature Integration (Case study: Provincial economic in Indonesia). *AIP Conference Proceedings*, 080022-1 - 080022-10. doi:<https://doi.org/10.1063/5.0106043>
- Ansari, R., Pohan, H. M., Elisa, E., Lubis, M., Sormin, A. S., Mora, J. L., & Harahap, H. J. (2022). Workshop Calistung Berbasis Pictures Themes bagi Anak Usia Dini. *Jurnal Ilmiah Kampus Mengajar*, 1-7.
- Apriliana, T., & Widodo, E. (2023). Analisis Cluster Hierarki untuk Pengelompokan Provinsi di Indonesia berdasarkan Jumlah Base Transceiver Station dan Kekuatan Sinyal. *Konvergensi Teknologi dan Sistem Informasi (KONSTELASI)*, 3, 286 - 296. doi:<https://doi.org/10.24002/konstelasi.v3i2.7143>
- Ayu, U., Hermawan, R., & Utami, R. D. (2021). Pendidikan sadar bencana melalui sosialisasi kebencanaan untuk meningkatkan kesiapsiagaan siswa MI Muhammadiyah Bulakrejo. *Jurnal Ilmiah Kampus Mengajar*, 1-11.
- Fauziyah, A. F., Amalia, N., & Kartikasari, E. D. (2022). Pengenalan Kebudayaan Indonesia melalui Boading Literasi SB Hulu Kelang Malaysia. *Buletin KKN Pendidikan*, 4(2), 161-166.
- Greene, W. H. (2012). *Econometric Analysis, Seventh Edition*. New Jersey: Prentice Hall.

- Gujarati, D. N. (2004). *Basic Econometrics, 4th Edition*. New York: McGraw-Hill Companies.
- Gundono. (2011). *Analisis Data Multivariat*. Yogyakarta: BPFE.
- Hafida, S. H. N., Ariandi, A. P., Ismiyatin, L., Wulandari, D. A., Reygina, N., Setyaningsih, T., ... & Amin, M. A. K. (2020). Pengenalan Etnobotani melalui Pembuatan Herbarium Kering di Lingkungan Sekolah MI Muhammadiyah Plumbon, Wonogiri. *Buletin KKN Pendidikan*, 2(2), 79-83.
- Hair, J. F., Anderson, R. E., Thatham, R. L., & Black, W. C. (2010). *Multivariate Data Analysis Seventh Edition*. New Jersey: Pearson Education, Inc.
- Imasdiani, Purnamasari, I., & Amijaya, F. D. (2022). Perbandingan Hasil Analisis Cluster dengan menggunakan Metode Average Linkage dan Metode Ward (Studi Kasus : Kemiskinan Di Provinsi Kalimantan Timur Tahun 2018). *Jurnal EKSPONENSIAL*, 13, 9-18. doi:https://doi.org/10.30872/ekspnensial.v13i1.875
- Irwan, Sanusi, W., & Hasanah, A. (2024). Perbandingan Analisis Cluster Metode Complete Linkage dan Metode Ward dalam Pengelompokkan Indeks Pembangunan Manusia di Sulawesi Selatan. *Journal of Mathematics, Computations, and Statistics*, 7, 75-86. doi:https://doi.org/10.35580/jmathcos.v7i1.2089
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis, 6th Edition*. New Jersey: Prentice Hall.
- Ma'aruf., Nugroho, S. P., & Veno, A. (2017). Pengukuran Daya Saing Klaster Batik, Konveksi dan Mebel Di Kabupaten Sragen. *Benefit: Jurnal Manajemen dan Bisnis*, 2(1), 62-77.
- Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, 04, 20-24. doi:https://doi.org/10.30591/jpit.v4i1.1253
- Odewumi, M. O., Falade, A. A., Adeniran, A. O., Akintola, D. A., Oputa, G. O., & Ogunlowo, S. A. (2019). Acquiring basic chemistry concepts through virtual learning in nigerian senior secondary schools. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 2(1), 56-67.
- Pradana, Y. A., Setyawati, Y., Dewi, L. P., Shobri, M. Q., Adhantoro, M. S., Kurniaji, G. T. B., & Romadloni, N. T. (2024). Penentuan Rute Optimal Wisata di Kota dan Kabupaten Madiun Menggunakan Algoritma Genetika. *Jurnal Keilmuan dan Keislaman*, 49-56.
- Saraswati, A. M., & Nugroho, A. W. (2021). Perencanaan Keuangan dan Pengelolaan Keuangan Generasi Z di Masa Pandemi Covid 19 melalui

- Penguatan Literasi Keuangan. *Warta LPM*, 24(2), 309-318.
- Setiawan, F. A., Arisanty, D., Hastuti, K. P., & Rahman, A. M. (2020). The Effect of Metacognitive Ability on Learning Outcomes of Geography Education Students. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 2(2), 82-90.
- Singh, A., Rana, A., & Yadav, A. (2013). K-means with Three Different Distances Metrics. *International Journal of Computer Applications*, 67, 13-17.
- Suleiman, Y., Hanafi, Z., & Muhajir, T. (2019). Influence of extracurricular services on students' academic achievement in secondary schools in Kwara State: A qualitative approach. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 1(2), 1-19.
- Supranto, J. (2010). *Analisis Multivariat, Arti dan Interpretasi*. Jakarta: Rineka Cipta.
- Suprpto, B., Simanjuntak, H., & Sulasminarti. (2021). Perbandingan Metode Nearest Neighbor, Ward, dan K-Means dalam Menentukan Cluster Data Kinerja Kantor Unit Bank ABC. *Jurnal Informasi dan Komputer*, 9, 53-65.
- doi:<https://doi.org/10.35959/jik.v9i1.199>
- Talakua, M. W., Leleury, Z. A., & Talluta, A. W. (2017). Analisis Cluster dengan Menggunakan Metode K-Means untuk Mengelompokkan Kab/Kota di Provinsi Maluku Berdasarkan IPM Tahun 2014. *Barekeng : Ilmu Matematika dan Terapan*, 119-128.
- Utami, R. D., Alfalah, Z. A. A., Prayitno, H. J., Desstya, A., Kartini, N. H., & Negara, S. P. P. S. (2024, July). Thematic Learning Innovation Approaching SCI (Smart, Creative, Innovative) Plus Elementary Schools in the Covid-19 Pandemic Era. In *International Conference on Education for All* (Vol. 2, No. 1, pp. 30-45).
- Vandevoorde, L. (2020). *Semantic Differences in Translation: Exploring the field of inchoativity (translation and Multilingual Natural Language Processing 13)*. Berlin: Language Science Press.
- Yamin, M., Saputra, A., & Deswila, N. (2020). Enhancing Critical Thinking in Analyzing Short Story "The Lazy Jack" Viewed from Identity Theory. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 3(1), 30-39.