SKRIPSI

ANALISIS *AGGLOMERATIVE HIERARCHICAL CLUSTERING* TINGKAT PENGANGGURAN TERBUKA DI INDONESIA MENGGUNAKAN JARAK *DYNAMIC TIME WARPING*

Diajukan sebagai salah satu syarat memperoleh gelar sarjana pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Sulawesi Barat

JUMARI
E0221529

PROGRAM STUDI STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS SULAWESI BARAT

TAHUN 2025

SURAT PERNYATAAN

Yang bertanda tangan dibawa ini:

Nama	: Jumari
Tempat/Tgl Lahir	: Riso 12 Juni 2001
NIM	: E0221529
Program Studi	: Statistika

Menyatakan bahwa tugas akhir dengan judul "Analisis *Agglomerative Hierarchical Clustering* Tingkat Pengangguran Terbuka di Indonesia Menggunakan Jarak *Dynamic Time Warping*" disusun berdasarkan prosedur ilmiah yang telah melalui pembimbingan dan bukan merupakan plagiat dari karya ilmiah/naskah yang lain. Apabilah di kemudian hari terbukti bahwa pernyataan ini tidak benar, maka saya bersedia menerima sanksi sesuai yang berlaku.

Majene, 14 Oktober 2025

yang menyatakan,

Jumari

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Jumari

Nim : E0221529

Judul : Analisis *Agglomerative Hierarchical Clustering* Tingkat Pengangguran Terbuka di Indonesia Menggunakan Jarak *Dynamic Time Warping*.

Telah berhasil dipertahankan di depan Tim Penguji (SK Nomor: 16/UN55.7/HK.04/2025 tanggal 26 Februari 2025) dan diterima sebagai sebagai bagian persyaratan memperoleh gelar Sarjana S.Stat pada Porgam Studi Statistika Fakultas Matematika dan Ilmu pengetahuan Alam Universitas Sulawesi Barat.

Disahkan Oleh:
Dekan FMIPA
Universitas Sulawesi Barat

Musafira, S.Si., M.Sc.
NIP. 197709112006604220002

Tim Penguji:

Ketua penguji : Musafira, S.Si., M.Sc. (.....)

Sekertaris : Muh. Hijrah, S.Pd., M.Si. (.....)

Pembimbing 1 : Muh. Hijrah, S.Pd., M.Si. (.....)

Pembimbing 2 : Muhammad Hidayatullah, S.Pd., M.Kom. (.....)

Penguji 1 : Reski Wahyu Yanti, S.Si., M.Si. (.....)

Penguji 2 : Retno Mayapada, S.Si., M.Si. (.....)

Penguji 3 : Putri Indi Rahayu, S.Si., M.Stat. (.....)

KATA PENGANTAR

Puji dan syukur yang mendalam penulis panjatkan kepada Tuhan yang Maha Esa, atas kasih karunia, penyertaan, kekuatan, dan anugerah-Nya yang luar biasa dalam setiap proses, tantangan, dan perjalanan penyusunan tugas akhir ini yang berjudul berjudul **"Analisis *Agglomerative Hierarchical Clustering* Tingkat Pengangguran Terbuka di Indonesia Menggunakan Jarak *Dynamic Time Warping*"** tepat pada waktunya. Disusun untuk memenuhi salah satu syarat dalam menyelesaikan Strata Satu (S1) pada Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sulawesi Barat.

Banyak hambatan yang penulis temui dalam penyusunan tugas akhir ini, namun berkat kerja keras, tekad yang kuat, serta bimbingan dan bantuan dari berbagai pihak yang turut memberikan andil baik secara langsung maupun tidak langsung, secara moril maupun materil terutama kepada kedua orang tua tercinta, Thomas Rombe dan Napa, para inspirasi hidup yang telah membagikan kasih tanpa pamrih kepada anak-anaknya, sehingga tugas akhir ini dapat diselesaikan dengan baik; untuk itu, penulis menyampaikan terima kasih dengan tulus dan penuh kerendahan hati kepada yang terhormat:

1. Tuhan Yesus Kristus, atas kasih yang tak terbatas, penyertaan yang setia, serta kekuatan dan pengharapan yang senantiasa engkau limpahkan dalam setiap langkah hidupku, khususnya selama proses panjang, penuh tantangan, dan pembelajaran dalam penyusunan tugas akhir ini. Tanpa campur tangan-Mu, karya ini tidak akan dapat terselesaikan dengan baik.
2. Bapak Prof. Dr. Muhammad Abdy, S.Si., M.Si., selaku Rektor Universitas Sulawesi Barat yang telah memberikan kesempatan kepada penulis untuk menempuh dan menyelesaikan pendidikan Strata Satu (S1) di Universitas Sulawesi Barat.
3. Ibu Musafira, S.Si., M.Sc., yang menjabat sebagai Pelaksana Tugas Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sulawesi Barat.

4. Bapak Hirman Rachman, S.Si., M.Si., selaku Wakil Dekan I pada Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sulawesi Barat.
5. Bapak Muh. Hijrah, S.Pd., M.Si., yang menjabat sebagai Ketua Koordinator Program Studi Statistika dan sekaligus sebagai Dosen Pembimbing I, yang telah dengan penuh kesabaran memberikan bimbingan, arahan, serta dukungan selama proses penyusunan tugas akhir ini hingga dapat terselesaikan dengan baik.
6. Bapak Muhammad Hidayatullah, S.Pd., M.Kom., selaku Pembimbing II, yang telah dengan penuh kesabaran memberikan bimbingan, arahan, serta dukungan selama proses penyusunan tugas akhir ini hingga dapat terselesaikan dengan baik.
7. Seluruh Dosen Penguji yang telah dengan penuh kesediaan meluangkan waktu untuk menguji serta memberikan saran dan kritik yang membangun demi penyempurnaan penyusunan skripsi ini.
8. Bapak/Ibu Dosen Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sulawesi Barat, yang telah membagikan ilmu pengetahuan dan wawasan berharga selama masa perkuliahan.
9. Seluruh staf Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sulawesi Barat, yang telah memberikan bantuan dalam berbagai urusan administrasi maupun kebutuhan lainnya selama proses studi.
10. Ucapan terima kasih saya sampaikan kepada seluruh keluarga besar Bapak dan ibu atas doa, semangat, serta nasihat yang senantiasa menguatkan sepanjang perjalanan ini.
11. Sahabat tercinta Satriwanti dan Kartini, yang telah setia menemani selama perkuliahan dan dalam berbagai suka duka, serta memberikan dukungan dan bantuan yang sangat berarti dalam proses penyelesaian skripsi ini.
12. Teman-teman kost tercinta Srirahmayani, Ceriayusnita, Rosnawati, Mutiara, Ispamuliati dan Didin Tolla. Kehangatan kebersamaan, tawa, serta dukungan kalian telah menjadi penyemangat selama masa studi dan penyusunan skripsi

ini.

13. Teman-teman Kelas Statistika A tercinta, Terima kasih atas kebersamaan, semangat, dan dukungan selama masa perkuliahan. Serta kerja sama kalian menjadi bagian berharga dalam perjalanan studi ini.

Kepada semua pihak yang tidak dapat saya sebutkan satu per satu, saya menyampaikan rasa terima kasih yang sedalam-dalamnya atas segala doa, dukungan, bantuan, dan motivasi yang telah diberikan selama proses penyusunan karya ini. Setiap bentuk perhatian, semangat, serta dorongan yang diberikan, baik secara langsung maupun tidak langsung, telah menjadi sumber kekuatan dan inspirasi bagi saya untuk terus berusaha hingga karya ini dapat terselesaikan dengan baik.

Majene, 15 Oktober 2025

Jumari
E221529)

DAFTAR SIMBOL

$d_{(uv)w}$: jarak DTW antara *cluster* pengnggabungan objek *u* dan *v* dengan objek *w*

$d_{(ij)k}$: jarak DTW antara *ij* dan *cluster k*

d_{ab} : jarak DTW antara objek *a* pada *cluster ij* dan objek *b* pada *cluster k*

N_{ij} : jumlah anggota pada *cluster ij*

N_k : jumlah anggota pada *cluster k*

w : jalur *warping* yaitu urutan pasangan titik (*i,j*) yang mencocokkan elemen dari deret *a d*an *b*

p : panjang jalur *warping w* yaitu jumlah pasangan titik (*i,j*) dalam jalur tersebut

$\delta_{(w_k)}$: jarak DTW antara elemen s_{i_k} dari deret *S* dan t_{j_k} dari deret *T* pada pasangan ke-*k*

r_{coph} : koefisien korelasi *cophenetic*

$\bar{d}$: rata - rata jarak DTW antara objek ke - *i* dan ke – *j*

d_{cij} : jarak *cophenetic* antara objek *i* dan *j*

$\bar{d}_c$: rata- rata jarak *cophenetic* antara objek *i* dan *j*

Conn (C) : indeks *connecticity*

$n_{i(j)}$: pengamatan tetangga terdekat dari data ke-*i* data ke-*j*

N : banyak pengamatan

L : banyak *cluster*

d_{min} : jarak terkecil objek pada *cluster* yang berbeda.

d_{max} : jarak terbesar antar objek pada *cluster* yang sama

a(i) : rata-rata jarak antara objek *i* dan objek lain dalam *cluster* yang sama

b(i) : rata-rata jarak antara *i* dan objek dalam *cluster* terdekat lainnya

ABSTRAK

Tingkat Pengangguran Terbuka (TPT) merupakan indikator penting dalam menilai kondisi ketenagakerjaan di Indonesia. Penelitian ini bertujuan mengelompokkan 34 provinsi berdasarkan pola fluktuasi TPT periode 2015–2024 menggunakan metode *aglomerative hierarchical clustering*. Hasil analisis menunjukkan bahwa Bali hampir selalu memiliki TPT terendah karena pariwisata padat karya, sedangkan TPT tertinggi bervariasi pada provinsi seperti Kepulauan Riau, Jawa Barat, DKI Jakarta, dan Kalimantan Timur akibat faktor industrialisasi, urbanisasi, serta ketergantungan pada sektor pertambangan. Evaluasi metode menunjukkan *average linkage* paling optimal dengan nilai koefisien korelasi *cophenetic* 0,722. *Clustering* menghasilkan dua kelompok, yaitu *cluster* 1 dengan rata-rata TPT tinggi dan cenderung bertahan, serta *cluster* 2 dengan rata-rata TPT lebih rendah dan relatif stabil. Temuan ini menegaskan adanya disparitas ketenagakerjaan antarprovinsi yang penting diperhatikan dalam kebijakan pengangguran di Indonesia.

Kata Kunci: TPT, *Agglomerative Hierarchical Clustering, Average Linkage,* Indonesia

ABSTRACT

The Open Unemployment Rate (OUR) is an important indicator for assessing labor market conditions in Indonesia. This study aims to cluster 34 provinces based on the fluctuation patterns of OUR from 2015 to 2024 using the agglomerative hierarchical clustering method. The results show that Bali almost consistently records the lowest OUR due to its labor-intensive tourism sector, while the highest rates vary among provinces such as Riau Islands, West Java, Jakarta, and East Kalimantan, driven by industrial fluctuations, large population size, rapid urbanization, and dependence on mining sectors. Method evaluation indicates that the average linkage approach is the most optimal, with the highest cophenetic correlation coefficient of 0.722. The clustering analysis produces two groups: cluster 1 with relatively high and persistent OUR, and cluster 2 with lower and more stable OUR. These findings highlight labor market disparities across provinces that need to be addressed in Indonesia's employment policies.

Keywords: *OUR, Agglomerative Hierarchical Clustering, Average Linkage, Indonesia*

BAB I
PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan negara dengan jumlah penduduk terbesar di kawasan Asia Tenggara, yang menjadi potensi besar dalam mendukung pembangunan nasional (Marliana, 2022). Indonesia mulai memasuki bonus demografi pada tahun 2012, sementara periode puncaknya diperkirakan terjadi pada tahun 2020 hingga 2035 (BPS, 2023). Bonus demografi merupakan kondisi ketika jumlah penduduk usia produktif 15 sampai 64 tahun berada dalam proporsi yang sangat besar. Keadaan ini memberikan peluang besar bagi pertumbuhan ekonomi karena ketersedian sumber daya manusia yang melimpah dan siap bekerja. Ketika ketersediaan lapangan kerja tidak mampu menyerap seluruh tenaga kerja, bonus demografi justru bisa memicu persoalan baru, seperti meningkatnya angka pengangguran (Putri & Suhartini, 2024).

Pengangguran merupakan masalah ekonomi tingkat makro yang berdampak langsung pada kelangsungan hidup manusia. Pengangguran dibedakan ke dalam dua kategori utama, yaitu pengangguran terselubung dan pengangguran terbuka (Ardian dkk., 2022). Pengangguran terselubung adalah tenaga kerja yang bekerja kurang dari 35 jam perminggu (Maddarangang, 2022). Pengangguran terbuka merupakan angkatan kerja yang tidak bekerja tetapi sedang aktif mencari pekerjaan (Padang & Murtala, 2020). Penting untuk meninjau lebih jauh mengenai tingkat pengangguran terbuka, yaitu indikator yang menggambarkan proporsi pengangguran terbuka terhadap total angkatan kerja dalam suatu wilayah dan periode waktu tertentu. Permasalahan TPT masih menjadi tantangan serius dalam pembangunan ketenagakerjaan di Indonesia. Karena tidak seimbangannya antara jumlah pencari kerja dan adanya lapangan kerja, ketidaksesuaian keterampilan dengan kebutuhan industri, serta dampak pandemi COVID-19 yang melanda Indonesia pada tahun 2020 menyebabkan banyak tenaga kerja kehilangan pekerjaan yang juga menjadi faktor tingginya TPT (Darmawan & Mifrahi, 2022).

TPT merupakan faktor penting yang dapat memengaruhi kemiskinan, sehingga dibutuhkan strategi yang tepat untuk memahami dinamika pengangguran di

Indonesia, salah satu strategi dengan mengelompokkan wilayah yang memiliki pola fluktuasi TPT yang serupa ke dalam beberapa *cluster*. Oleh karena itu, untuk mengetahui pola TPT yang dominan serta merumuskan kebijakan ketenagakerjaan yang lebih tepat sasaran, dapat dilakukan analisis *clustering* terhadap provinsi-provinsi di Indonesia berdasarkan kemiripan pola perubahan TPT dari tahun ke tahun (Nabilla & Primandari, 2024).

Clustering merupakan proses pembagian sekumpulan data menjadi kelompok-kelompok yang disebut *cluster*, dimana objek dalam satu *cluster* memiliki kemiripan karakteristik satu sama lain dan berbeda dengan *cluster* lainnya (Annisa dkk., 2022). Proses pengelompokan *cluster* juga dapat dilakukan melalui dua pendekatan, yaitu metode hierarki dan metode non-hierarki. Metode hierarki digunakan untuk membentuk struktur bertingkat dari kumpulan data dengan mendasarkan pada tingkat kesamaan karakteristik antar objek. Sedangkan non-hierarki mengelompokkan objek dengan jumlah *cluster* yang sudah ditentukan sebelumnya (Syafiyah dkk., 2022).

Dalam pendekatan *cluster* non-hierarki, *k-means*, *k-medoids* dan *fuzzy k-means* merupakan metode yang paling populer digunakan (Widyadhana dkk., 2021). Imasdiani dkk., (2022) menyatakan bahwa metode hierarki lebih efisien dibanding non-hierarki karena langsung membentuk dendogram atau tingkatan sendiri, sehingga memudahkan interpretasi tanpa perlu menentukan jumlah *cluster* di awal. Secara umum, metode hierarki dibagi menjadi dua yaitu metode *agglomerative* dan metode *divisive*, akan tetapi *agglomerative* merupakan metode hierarki yang paling sering digunakan. Metode *agglomerative* mencakup beberapa teknik penggabungan seperti *single linkage, complete linkage, average linkage,* median *linkage, ward* dan *centroid* (Damayanti & Wijayanto, 2021).

Agglomerative adalah metode atau algoritma dalam analisis *cluster* yang bertujuan untuk membentuk struktur *cluster* secara hierarki. Pendekatan ini diawali dengan menganggap setiap data sebagai satu *cluster* individu, kemudian secara bertahap menggabungkan *cluster-cluster* yang memiliki tingkat kemiripan tertinggi berdasarkan metrik atau jarak tertentu yang telah ditentukan sebelumnya (Sachrrial & Iskandar, 2023). Dalam proses penggabungan pada analisis *clustering*, diperlukan

ukuran kesamaan antar data. Umumnya, perhitungan jarak *Euclidean* sering digunakan, namun metode ini memiliki keterbatasan ketika diterapkan pada data deret waktu. Oleh karena itu, pada data deret waktu, pengukuran kesamaan antar data dilakukan menggunakan metode perhitungan jarak DTW yang mampu mengukur kemiripan pola antar data deret waktu meskipun terjadi pergeseran waktu atau kecepatan perubahan yang berbeda, sehingga lebih akurat dibandingkan metode jarak seperti *euclidean* dalam menganalisis perubahan pada data deret waktu (Nurul, 2024). Daulay dkk., (2023) menjelaskan bahwa metode DTW lebih efektif digunakan untuk mengatasi perbedaan waktu antara data uji dan data referensi karena mampu melakukan penyesuaian terhadap ketidaksesuaian dalam waktu.

Beberapa penelitian telah memanfaatkan metode analisis *clustering* dengan mengunakan jarak DTW, salah satunya dilakukan oleh Utami 2017 tentang perbandingan hasil pengelompokkan indikator kesehatan di Indonesia tahun 2015, dengan menggunakan perhitungan jarak DTW dan diperoleh metode terbaiknya yaitu *single linkage*. Kemudian pada tahun 2023 Tias dan Arum melakukan pengelompokan kecamatan di Kabupaten Merangin berdasarkan produksi tanaman perkebunan tahun 2021, dengan menggunakan metode *agglomerative hierarchical clustering* berdasarkan jarak DTW, sehingga diperoleh metode terbaiknya yaitu *average linkage*. Ditahun yang sama, Riani dan Sofra melakukan pengelompokan berdasarkan garis kemiskinan di Provinsi Jawa Barat dengan pendekatan *time series based clustering* menggunakan jarak DTW, dan diperoleh metode terbaiknya yaitu *complete linkage*.

Berdasarkan latar belakang diatas, penulis tertarik melakukan penelitian dengan menglompokan TPT di Indonesia pada periode tahun 2015 hingga 2024 mengunakan metode *single linkage*, *average linkage*, dan *complete linkage* dengan perhitungan jarak DTW. Penelitian ini diharapkan dapat mengidentifikasi pola-pola perubahan TPT antar provinsi secara lebih akurat.

1.2 Rumusan Masalah

Berdasarkan uraian pada rumusan masalah diatas, maka rumusan masalah untuk penelitian ini yaitu :

1. Bagaimana pola perkembangan TPT antar provinsi di Indonesia pada tahun 2015–2024?
2. Metode manakah yang paling efektif di antara *single linkage*, *average linkage*, dan c*omplete linkage* dalam pengelompokan Provinsi di Indonesia berdasarkan TPT tahun 2015 hingga 2024?
3. Bagaimana karakteristik provinsi-provinsi di Indonesia berdasarkan hasil pengelompokan TPT menggunakan metode terbaik?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan dari penelitian ini adalah:

1. Untuk menganalisis pola perkembangan TPT antar provinsi di Indonesia pada tahun 2015–2024.
2. Untuk mengevaluasi dan membandingkan efektivitas metode *single linkage*, *average linkage*, dan *complete linkage* dalam pengelompokan provinsi di Indonesia berdasarkan data TPT tahun 2015 hingga 2024.
3. Untuk mengetahui dan menganalisis karakteristik provinsi-provinsi di Indonesia berdasarkan hasil pengelompokan TPT dengan menggunakan metode *clustering* terbaik.

1.4 Manfaat Penelitian

Adapun manfaat yang dapat diambil dari penelitian ini adalah:

1. Untuk memberikan kontribusi metodologis melalui evaluasi efektivitas tiga metode *cluster single linkage, average linkage*, dan *complete linkage*, sehingga dapat membantu peneliti selanjutnya dalam memilih metode yang paling tepat untuk analisis data time series antar wilayah.
2. Untuk menyediakan informasi berbasis data mengenai karakteristik kelompok provinsi yang memiliki kemiripan perubahan TPT, yang bermanfaat bagi

pemerintah dan untuk strategi penanggulangan pengangguran secara lebih terarah dan efisien.

1.5 Batasan Masalah

Agar penelitian tugas akhir ini tidak keluar dari pokok permasalahan maka ruang lingkup pembahasan di batasi dengan masalah sebagai berikut:

1. Analisis dilakukan terhadap 34 provinsi di Indonesia, tanpa menyertakan beberapa provinsi baru hasil pemekaran di wilayah Papua.
2. Penelitian ini menggunakan tahun sebagai variabel untuk menganalisis pola perubahan TPT di setiap provinsi di Indonesia dalam rentang tahun 2015 hingga 2024.
3. Penelitian ini hanya berfokus pada analisis pola perubahan TPT antar provinsi, tanpa mengkaji faktor-faktor penyebab atau variabel lain yang memengaruhi TPT.

BAB II
TINJAUAN PUSTAKA

2.1 Tingkat Pengangguran Terbuka

Tingkat pengangguran terbuka (TPT) merupakan angka yang menggambarkan jumlah penduduk usia kerja yang sedang mencari pekerjaan, mempersiapkan usaha, merasa tidak memiliki peluang untuk mendapatkan pekerjaan, atau sudah memiliki pekerjaan tetapi belum mulai bekerja. TPT mengacu pada persentase jumlah pengangguran yang aktif mencari pekerjaan dibandingkan dengan total angkatan kerja (Wahyuni, 2021).

2.2 *Preprocessing* Data

*Preprocessin*g data dapat diartikan sebagai proses *cleaning* data yang bertujuan untuk menangani nilai-nilai yang hilang dalam dataset serta menghilangkan gangguan atau ketidak pastian yang terdapat pada data (Salsabila & Ridwan, 2024).

2.3 *Machine Learning*

Machine learning (ML) adalah salah satu cabang dari kecerdasan buatan (AI) yang berfokus pada penelitian, perancangan, dan pengembangan algoritma yang dapat mempelajari pola dari data dan membuat prediksi tanpa memerlukan pemrograman yang jelas. Pendekatan ML memiliki kemampuan untuk mengolah data dengan dimensi tinggi dan multivariat, serta dapat mengidentifikasi pola-pola mendalam dalam data, bahkan dalam lingkungan yang kompleks dan dinamis (Andriani, 2021). Dalam pembelajaran mesin, terdapat dua pendekatan utama, yaitu *supervised learning* dan *unsupervised learning* (Abijono dkk., 2021).

2.3.1 *Supervised Learning*

Supervised learning merupakan metode dimana sistem diberikan himpunan data latih yang berisi informasi tentang *input* dan *output* yang diharapkan. Dengan demikian, sistem dapat mempelajari hubungan dari data yang tersedia. Sistem akan mengidentifikasi pola dalam data tersebut, yang kemudian digunakan sebagai panduan untuk memproses kumpulan data berikutnya.

2.3.2 *Unsupervised Learning*

Unsupervised learning bersifat deskriptif yang berguna untuk mengelompokkan atau mengkategorikan data, dan tidak mendapat *training* data karena algoritma ini bukan bersifat prediktif, sehingga membutuhkan pembelajaran dari data yang telah ada.

2.4 Analisis *Cluster*

Analisis *cluster* merupakan salah satu teknik penting dalam analisis data statistik yang digunakan untuk mengelompokkan objek secara otomatis ke dalam beberapa kelompok *cluster* berdasarkan tingkat kemiripan karakteristiknya**.** Metode ini bertujuan untuk membentuk *cluster* yang memiliki tingkat homogenitas tinggi di dalam kelompok antar *cluster* serta tingkat heterogenitas tinggi antar kelompok antar *cluster* sehingga objek-objek dalam satu *cluster* memiliki kesamaan yang signifikan, sementara berbeda secara signifikan dengan objek di *cluster* lainnya. Analisis *cluster* dapat diterapkan pada data deret waktu, yang memerlukan prosedur dan algoritma pengelompokan yang berbeda dibandingkan dengan pengelompokan data potongan lintas waktu *cross-section*. Proses pembentukan *cluster* pada data deret waktu melibatkan prosedur dan algoritma yang disesuaikan, karena data ini terdiri dari serangkaian pengamatan yang terurut berdasarkan indeks waktu dengan interval waktu yang tetap (Dani dkk., 2020).

2.5 *Clustering Hierarchical*

Clustering hierarki terdiri dari dua pendekatan utama, yaitu metode *agglomerative* (penggabungan) dan *divisive* (pemisahan). Dalam pendekatan *agglomerative*, setiap objek atau identitas awalnya dianggap sebagai *cluster* tunggal, sehingga jumlah *cluster* sama dengan jumlah objek. Kemudian, dua *cluster* yang memiliki tingkat kemiripan paling tinggi akan digabung menjadi satu, sehingga jumlah *cluster* berkurang satu pada setiap tahapannya. Sebaliknya, metode *divisive* dimulai dari satu *cluster* besar yang mencakup seluruh data. Proses dilanjutkan dengan memisahkan data yang memiliki perbedaan paling mencolok ke dalam kelompok-kelompok yang lebih kecil. Tahapan pemisahan ini terus berlangsung hingga setiap data atau observasi membentuk *cluster* tersendiri. Seluruh proses ini

divisualisasikan dalam bentuk dendogram atau diagram pohon, yang menunjukkan hubungan dan tingkat kedekatan antar objek secara bertahap (Ulinnuha & Veriani, 2020).

Secara umum, algoritma *aggolmerative hierarchical clustering* untuk mengelompokkan *N* objek adalah sebagai berikut (Rachmatin, 2014).

1. Mulai dengan *N cluster*, setiap *cluster* mengandung unsur tunggal dan sebuah matriks simetris $D:\{d_{jl}\}$
2. Tentukan jarak terhadap pasangan *cluster* terdekat
3. Menggabungkan masing-masing *cluster* yang telah ditunjukkan memiliki jarak yang relatif dekat.
4. Ulangi langkah 2 dan 3 sebanyak $(n-1)$ kali, sampai semua objek akan berada dalam *cluster* tunggal.

2.5.1 *Agglomerative Hierarchical Clustering*

Agglomerative adalah salah satu pendekatan dalam metode *hierarchical clustering* yang di gunakan untuk pengelompokan dengan cara memulai dari masing-masing data individu sebagai *cluster* tersendiri, kemudian digabungkan secara bertahap. Pendekatan ini dikenal sebagai metode *bottom-up* (Azzahra & Wijayanto 2022). Langkah-langkah proses *clustering* menggunakan metode hierarki *agglomerative* (Suhirman & Wintolo 2019).

a. Menghitung matriks jarak antara masing-masing data.
b. Menggabungkan dua kelompok dengan jarak terdekat sesuai parameter kedekatan yang telah ditetapkan.
c. Memperbarui matriks jarak antar data untuk mencerminkan kedekatan antara kelompok baru dan kelompok lain yang tersisa.
d. Mengulangi langkah b dan c hingga terbentuk satu kelompok akhir.

2.5.2 *Single Linkage*

Single linkage merupakan metode yang mengelompokkan objek berdasarkan jarak terkecil antar objek (Tias & Arum 2023). Langkah-langkah penerapan metode *single linkage* adalah sebagai berikut (Pusdiktasari dkk., 2021).

a. Menghitung jarak DTW antara setiap pasangan sub-sampel.
b. Menentukan pasangan sub-sampel yang memiliki jarak terkecil dalam himpunan jarak $D = \{d_{(ij)}\}$.
c. Mengukur jarak antara *cluster* yang terbentuk dan objek lainnya yang belum tergabung.
d. Berdasarkan algoritma tersebut, jarak antara sub-sampel (*ij)* dan *cluster* lain *(k)* dihitung menggunakan rumus berikut:

$$d_{(UV)W} = min\ *d_{UV}, d_{VW}+ \quad (2.1)$$

dengan,

$d_{(uv)w}$: jarak antar *cluster* pengnggabungan objek *u* dan *v* dengan objek .

d_{uv} : jarak antar objek *u* dan objek *v*

d_{vw} : jarak antar objek *v* dan objek *w*

2.5.3 *Average Linkage*

Average linkage merupakan teknik pengelompokan di mana jarak antara dua *cluster* ditentukan berdasarkan rata-rata jarak dari seluruh pasangan data yang berasal dari masing-masing *cluster*. Dengan kata lain, metode ini menghitung jarak antar *cluster* dengan merata-ratakan jarak antar semua titik yang berada pada *cluster* satu dengan semua titik di *cluster* lainnya. Berikut rumus menghitung rata-rata jarak pada metode *average linkage* (Novitasari & Arofah 2023).

$$d_{(ij)K} = \frac{\sum_a \sum_b d_{ab}}{N_{ij} N_k} \quad (2.2)$$

dengan,

$d_{(ij)k}$: jarak antara *ij* dan *cluster k*

d_{ab} : jarak antara objek *a* pada *cluster ij* dan objek *b* pada *cluster k*

N_{ij} : jumlah anggota pada *cluster ij*

N_k : jumlah anggota pada *cluster k*

2.5.4 *Complete Linkage*

Complete linkage adalah metode pengelompokan objek berdasarkan jarak maksimum (terjauh) antar objek, kemudian proses pengelompokan dilanjutkan dengan memperhatikan jarak antar variabel yang semakin mendekat. Rumus dari *complete linkage* dapat dinyatakan sebagai berikut (Tias & Arum 2023).

$$d_{(UV)W} = max\ *d_{UV}, d_{VW}+ \quad (2.3)$$

dengan,

$d_{(uv)w}$: jarak antar *cluster* penggabungan objek *u dan* v dengan objek *w*

d_{uv} : jarak antar objek *u dan w*

d_{vw} : jarak antar objek *v dan w*

2.6 *Divisive*

Divisive merupakan kebalikan dari metode *agglomerative* dalam hierarki *cluster*. Dalam metode ini, proses dimulai dengan seluruh objek dalam satu *cluster*, kemudian *cluster* tersebut secara bertahap dibagi menjadi dua bagian dengan nilai terkecil pada setiap langkahnya. Proses ini terus berlanjut hingga setiap objek berada dalam *cluster* tersendiri. Dengan demikian, teknik *divisive* dapat disimpulkan sebagai pendekatan hierarki yang memulai pengelompokan dari satu *cluster* besar, kemudian membaginya secara bertahap (awal menjadi dua *cluster*, lalu tiga, dan seterusnya) hingga setiap objek menjadi *cluster* tunggal (Tarigan, 2024).

2.7 *Time Series*

Menurut Arumsari & Dani (2021), *time series* atau runtun waktu adalah sekumpulan observasi data yang terurut berdasarkan waktu dengan interval yang konsisten. Data yang tersusun secara berurutan dengan interval tetap, seperti mingguan, bulanan, atau tahunan, termasuk dalam lingkup metode *time series***.** Data *time series* ini diorganisir berdasarkan urutan waktu yang konsisten. Sementara itu, analisis runtun waktu merujuk pada pengolahan data untuk memahami atau

meramalkan kondisi di masa depan Secara umum terdapat empat pola utama dalam data *time series*.

1. Pola horizontal terjadi ketika data berfluktuasi secara acak tanpa menunjukkan tren yang jelas, tetapi dapat dipengaruhi oleh kejadian tidak terduga.
2. Pola tren menggambarkan kecenderungan data yang bergerak naik atau turun dalam jangka panjang.
3. Pola musiman menunjukkan fluktuasi data yang terjadi secara periodik dalam satu tahun, seperti triwulanan, kuartalan, bulanan, mingguan, atau harian.
4. Pola siklis mencerminkan fluktuasi data dalam periode lebih dari satu tahun, yang biasanya berkaitan dengan siklus ekonomi atau tren jangka panjang lainnya.

2.8 Jarak *Dynamic Time Warping*

Dynamic time warping (DTW) merupakan sebuah algoritma yang digunakan untuk membandingkan dua deret waktu serta menentukan lintasan optimal yang menghasilkan jarak minimum di antara keduanya. Tidak seperti algoritma tradisional yang membandingkan urutan diskrit atau nilai kontinu secara langsung, DTW menawarkan pendekatan yang lebih fleksibel. Secara matematis, metode DTW dapat dirumuskan sebagai berikut (Amatullah dkk., 2025).

$$DTW(S,T) = \min_{w} \left\lceil \sum_{k=1}^{p} \delta(w_k) \right\rceil \tag{2.4}$$

dengan,

S,T : dua deret waktu *(time series)* yang digunakan sebagai objek

w : jalur *warping* yaitu urutan pasangan titik (i,j) yang mencocokkan elemen dari deret S *d*an T

p : panjang jalur *warping* w yaitu jumlah pasangan titik (i,j) dalam jalur tersebut

$\delta_{(w_k)}$: jarak antara elemen s_{i_k} dari deret S dan t_{j_k} dari deret T pada pasangan ke-k

Untuk memperoleh jarak DTW antara dua deret waktu, dilakukan beberapa langkah sebagai berikut:

1. Langkah pertama adalah menyiapkan dua deret waktu dalam rentang waktu tertentu. Deret waktu pertama dinyatakan sebagai $A = [a_1, a_2, \ldots . , a_n]$, dan deret waktu kedua sebagai $B = [b_1, b_2, . \ldots , b_n]$.
2. Dilakukan perhitungan matriks jarak antara elemen-elemen dari kedua deret waktu dengan menggunakan rumus $D(i,j) = |a_i - b_j|$, yaitu selisih absolut antara dua titik pada waktu ke-*i* dan *j*.
3. Setelah diperoleh matriks jarak awal, selanjutnya dibentuk matriks jarak kumulatif secara bertahap. Setiap elemen diisi dengan penjumlahan nilai jarak saat ini dan nilai minimum dari tiga arah sebelumnya (atas, kiri, dan diagonal), untuk membentuk jalur kumulatif dengan biaya total terkecil, yang dirumuskan sebagai:

   $$D(i,j) = d(i,j) + \min \{D(i-1,j), D(I,j-1), D(i-1,j-1).$$

   Dengan $D(i,j)$ adalah elemen matriks jarak kumulatif, dan $d(i,j)$ adalah elemen dari matriks jarak awal.
4. Jarak DTW diperoleh dari nilai terkecil pada akhir jalur optimal (*minimum warping path*) yang terbentuk dalam matriks kumulatif.

2.9 Koefisien Korelasi *Cophenetic*

Setelah dilakukan analisis *clustering* menggunakan metode dalam *agglomerative hierarchical clustering*, tahap selanjutnya adalah melakukan uji validasi untuk menentukan metode *clustering* yang menghasilkan struktur *cluster* paling optimal. Proses validasi ini menggunakan koefisien korelasi *cophenetic*, yaitu ukuran korelasi antara nilai-nilai ketidakmiripan asli (dalam hal ini berdasarkan matriks jarak, seperti *Euclidean* atau DTW) dengan nilai-nilai ketidakmiripan yang direpresentasikan oleh dendogram dalam bentuk matriks *cophenetic*. Rumus perhitungan koefisien korelasi *cophenetic* didefinisikan sebagai berikut (Septianingsih, 2022).

$$r_{coph} = \frac{\sum_{i<j}(d_{ij} - \bar{d})(d_{cij} - d_c)}{\sqrt{\left[\sum_{i<j}(d_{ij} - \bar{d})^2\right]\left[\sum_{i<j}(d_{cij} - \bar{d}_c)^2\right]}} \quad (2.5)$$

dengan,

r_{coph} : koefisien korelasi *cophenetic*

d_{ij} : jarak DTW objek ke - *i* dan ke - *j*

$\bar{d}$: rata - rata jarak DTW antara objek ke - *i* dan ke – *j*

d_{cij} : jarak *cophenetic* antara objek *i* dan *j*

$\bar{d}_c$: rata- rata jarak *cophenetic* antara objek *i* dan *j*

Nilai koefisien korelasi *cophenetic* umumnya berada dalam rentang -1 hingga 1. Semakin mendekati angka 1, nilai tersebut menunjukkan bahwa hasil analisis *cluster* semakin representatif dan memberikan solusi *cluster* yang lebih baik.

2.10 Validitas *Cluster*

Validasi *cluster* yang bertujuan untuk menentukan jumlah *cluster* yang optimal terbagi menjadi tiga jenis, yaitu validasi internal, validasi stabilitas, dan validasi biologis. Pada penelitian ini, metode validasi yang digunakan adalah validasi internal dan validasi stabilitas untuk menilai keakuratan hasil pengelompokan (Afira & Wijayanto, 2021).

2.10.1 Indeks *Connectivity*

Indeks *connectivity* berada pada rentang nol (0) sampai tak terhingga (∞). Semakin rendah nilai *connectivity* maka *cluster* yang terbentuk semakin baik. Adapun rumus yang digunakan untuk menghitung *connectivity* yaitu (Suraya & Wijayanto, 2022).

$$Conn(C) = \sum_{i=1}^{N}\sum_{j=1}^{L} X_i \quad (2.6)$$

dengan,

Conn (C) : indeks *connecticity*

$nn_{i(j)}$: pengamatan tetangga terdekat dari data ke-*i* data ke-*j*

N : banyak pengamatan

L : banyak *cluster*

2.10.2 Indeks *Dunn*

Indeks *dunn* merupakan rasio antara jarak minimum antar objek dari *cluster* yang berbeda dengan jarak maksimum antar objek dalam satu *cluster*. Nilai indeks *dunn* yang lebih besar menunjukkan bahwa jumlah *cluster* yang terbentuk semakin optimal. Indeks ini dilambangkan dengan huruf D dan dihitung menggunakan rumus pada persamaan berikut (Septianingsih, 2022).

$$D = \frac{d_{\min}}{d_{\max}} \tag{2.7}$$

dengan,

d_{min} : jarak terkecil objek pada *cluster* yang berbeda

d_{max} : jarak terbesar antar objek pada *cluster* yang sama

2.10.3 Indeks *Silhouette*

Indeks *silhouette* digunakan untuk mengukur tingkat keseragaman anggota dalam satu *cluster* (homogenitas internal) serta perbedaan antar anggota dari *cluster* yang berbeda (heterogenitas eksternal). Nilai koefisien ini berada pada rentang -1 hingga 1. Semakin mendekati angka 1, maka semakin baik hasil validasi terhadap jumlah *cluster* yang terbentuk. Sebaliknya, jika nilai mendekati -1, maka menunjukkan bahwa pemisahan *cluster* kurang optimal (Septianingsih 2022). Nilai koefisien *silhouette* dihitung menggunakan rumus berikut (Mahadesyawardani dkk., 2024).

$$s(i) = \frac{b(a) - a(i)}{\max\left(a(i), b(i)\right)} \tag{2.8}$$

dengan,

a(i) : rata-rata jarak antara objek *i* dan objek lain dalam *cluster* yang sama

b(i) :rata-rata jarak antara *i* dan objek dalam *cluster* terdekat lainnya

BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan hasil analisis *cluster* TPT di Indonesia, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Pola perkembangan TPT di Indonesia pada periode 2015–2024 secara umum menunjukkan tren yang fluktuatif dan tidak konsisten, di mana posisi tertinggi bergantian ditempati oleh beberapa provinsi seperti Kepulauan Riau, Kalimantan Timur, DKI Jakarta, dan Jawa Barat, sedangkan Bali secara dominan menempati posisi terendah hampir di seluruh tahun pengamatan.
2. Metode *average linkage* terbukti menjadi pendekatan *agglomerative hierarchical clustering* yang paling optimal dalam mengelompokkan provinsi di Indonesia berdasarkan data TPT tahun 2015–2024. Hal ini didasarkan pada nilai koefisien korelasi *cophenetic* tertinggi sebesar 0,722, dibandingkan dengan *single linkage* 0,552 dan *complete linkage* 0,691, yang menunjukkan bahwa *average linkage* paling baik merepresentasikan pola struktur data.
3. Jumlah *cluster* optimal yang terbentuk adalah dua, masing-masing dengan karakteristik berbeda. *Cluster* 1 terdiri dari provinsi-provinsi dengan rata-rata TPT yang tinggi dan cenderung bertahan dari tahun 2012 hingga 2024. Sementara itu, *cluster* 2 menunjukkan rata-rata TPT yang lebih rendah secara konsisten dan mengindikasikan kondisi ketenagakerjaan yang lebih baik.

5.2 Saran

1. Peneliti berikutnya juga dapat memperdalam analisis terhadap masing-masing *cluster*, terutama dengan meninjau lebih lanjut faktor-faktor penyebab tingginya TPT di *cluster* 1, serta merancang kebijakan atau rekomendasi berdasarkan stabilitas ketenagakerjaan di *cluster* 2.
2. Penelitian mendatang dapat mempertimbangkan penggunaan metode *clustering* alternatif seperti *K-Means*, DBSCAN, atau pendekatan spasial, serta membandingkannya dengan hasil dari metode *agglomerative*

hierarchical clustering, untuk melihat konsistensi dan keakuratan pengelompokan wilayah berdasarkan pola perubahan TPT.

DAFTAR PUSTAKA

Abijono, H., Santoso, P., & Anggreini, N. L. (2021). Algoritma Supervised Learning Dan Unsupervised Learning Dalam Pengolahan Data. *Jurnal Teknologi Terapan: G-Tech*, *4*(2), 315–318. https://doi.org/10.33379/gtech.v4i2.635

Afira, N., & Wijayanto, A. W. (2021). Analisis Cluster dengan Metode Partitioning dan Hierarki pada Data Informasi Kemiskinan Provinsi di Indonesia Tahun 2019. *Komputika : Jurnal Sistem Komputer*, *10*(2), 101–109. https://doi.org/10.34010/komputika.v10i2.4317

Amatullah, F. F., Ilmani, E. A., Fitrianto, A., Erfiani, E., & Jumansyah, L. M. R. D. (2025). Clustering Time Series Forecasting Model for Grouping Provinces in Indonesia Based on Granulated Sugar Prices. *Journal of Applied Informatics and Computing*, *9*(1), 121–130. https://doi.org/10.30871/jaic.v9i1.8840

Andriani, A. Z. (2021). *Predictive maintenance (PdM)* menggunakan active dan semi-*supervised machine learning* pada mesin industri (Doctoral dissertation, Institut Teknologi Sepuluh Nopember).

Annisa, K., Ginting, B. S., & Syar, M. A. (2022). Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering Pada Pdam Langkat. *Jurnal Sistem Informasi Kaputama (JSIK)*, *6*(2), 165–179. https://doi.org/10.59697/jsik.v6i2.167

Ardian, R., Syahputra, M., & Desmawan, D. (2022). Pengaruh Pertumbuhan Ekonomi Terhadap Tingkat Pengangguran Terbuka Di Indonesia. *Jurnal Ekonomi, Bisnis Dan Manajemen*, *1*(3), 190–198. https://doi.org/10.58192/ebismen.v1i3.90

Arumsari, M., & Dani, A. (2021). Peramalan Data Runtun Waktu menggunakan Model Hybrid Time Series Regression – Autoregressive Integrated Moving Average. *Jurnal Siger Matematika*, *2*(1), 1–12. https://doi.org/10.23960/jsm.v2i1.2736

Azzahra, A., & Wijayanto, A. W. (2022). Perbandingan Agglomerative Hierarchical dan K-Means Dalam Pengelompokkan Provinsi Berdasarkan Pelayanan Kesehatan Maternal. *SISTEMASI: Jurnal Sistem Informasi*, *11*(2), 481–495. http://download.garuda.kemdikbud.go.id/article.php?article=2929100%5C&val=10006%5C&title=Comparison of Agglomerative Hierarchical and K-Means in Grouping Provinces Based on Maternal Health Services

Badan Pusat Statistik (BPS) Indonesia. (2012–2024). *Tingkat pengangguran terbuka*. https://www.bps.go.id/id/statistics-table/2/NTQzIzI=/unemployment-rate-by-province.html (diakses 28 November 2024).

Badan Pusat Statistik. (2023). Bonus demografi dan Visi Indonesia Emas 2045. https://bigdata.bps.go.id/documents/datain/2023(diakses 30 Mei 2025)

Damayanti, A. R., & Wijayanto, A. W. (2021). Comparison of Hierarchical and Non-Hierarchical Methods in Clustering Cities in Java Island using the Human Development Index Indicators year 2018. *Eigen Mathematics Journal*, *4*(1), 8–17. https://doi.org/10.29303/emj.v4i1.89

Dani, A. T. R., Wahyuningsih, S., & Rizki, N. A. (2020). Pengelompokkan Data Runtun Waktu menggunakan Analisis Cluster (Studi Kasus: Nilai Ekspor Komoditi Migas dan Nonmigas Provinsi Kalimantan Timur Periode Januari 2000-Desember 2016). *Jurnal EKSPONENSIAL*, *11*(1), 29–38.

Darmawan, A. S., & Mifrahi, M. N. (2022). Analisis Tingkat Pengangguran Terbuka di Indonesia Periode Sebelum dan Saat Pandemi Covid-19. *Jurnal Kebijakan Ekonomi Dan Keuangan*, *1*(1), 111–118.

Daulay, R. A., Ikhsan, M., & Hasugian, A. H. (2023). Implementasi metode dynamic time warping pada aplikasi kamus bahasa isyarat Indonesia sebagai media komunikasi. *Jurnal Ilmiah Binary STMIK Bina Nusantara Jaya Lubuklinggau*, 5(2), 160–166.

Imasdiani, I., Purnamasari, I., & Amijaya, F. D. T. (2022). Perbandingan Hasil Analisis Cluster Dengan Menggunakan Metode Average Linkage Dan Metode Ward. *Eksponensial*, *13*(1), 9. https://doi.org/10.30872/eksponensial.v13i1.875

Maddarangang, F. R. (2022). analisis determinan pengangguran terselubung di indonesia (Doctoral dissertation, Universitas Hasanuddin).

Mahadesyawardani, A., Zhafirab, A. A., Ariyawan, J., Humaira, E. P., Mardianto, M. F. F., Amelia, D., & Ana, E. (2024). Pengelompokan Kabupaten dan Kota di Jawa Timur berdasarkan Percepatan Pemulihan Ekonomi Menggunakan

Marliana, L. (2022). Analisis pengaruh indeks pembangunan manusia, pertumbuhan ekonomi dan upah minimum terhadap tingkat pengangguran terbuka di Indonesia. *Ekonomis: Journal of Economics and Business*, 6(1), 87–91.

Nabilla, W. B., & Primandari, A. H. (2024). Analisis Clustering Tingkat Pengangguran Terbuka di Provinsi DIY Tahun 2010-2022 dengan Dynamic Time Warping. *Emerging Statistics and Data Science Journal*, *2*(1), 135–144. https://doi.org/10.20885/esds.vol2.iss.1.art13

Novitasari, P., & Arofah, I. (2023). Penerapan Metode Clustering Average Linkage Untuk Mengelompokkan Kabupaten/Kota Di Provinsi Sumatera Utara Berdasarkan Indikator Kemiskinan. *MathVision : Jurnal Matematika*, *5*(1), 22–27. https://doi.org/10.55719/mv.v5i1.604

Nurul L. (2024). Perbandingan Kinerja K-Means Dan K-Medoids Pada Time Series-Based Clustering.

Padang, L., & Murtala, M. (2020). Pengaruh Jumlah Penduduk Miskin Dan Tingkat Pengangguran Terbuka Terhadap Pertumbuhan Ekonomi Di Indonesia. *Jurnal Ekonomika Indonesia*, *9*(1), 9. https://doi.org/10.29103/ekonomika.v9i1.3167

Pusdiktasari, Z. F., Sasmita, W. G., Fitrilia, W. R., Fitriani, R., & Astutik, S. (2021). The Clustering of Provinces in Indonesia by The Economic Impact of Covid-19 using Cluster Analysis. *Indonesian Journal of Statistics and Its Applications*, *5*(1), 117–129. https://doi.org/10.29244/ijsa.v5i1p117-129

Putri, F. R. G., & Suhartini, A. M. (2024). Pengaruh Pengaruh Bonus Demografi Dan Industrialisasi Terhadap Pengangguran Terdidik Usia Muda Di Indonesia Tahun 2018-2022. *Seminar Nasional Official Statistics*, *2024*(1), 269–278. https://doi.org/10.34123/semnasoffstat.v2024i1.2152

Rachmatin, D. (2014). Aplikasi Metode-Metode Agglomerative Dalam Analisis Klaster Pada Data Tingkat Polusi Udara. In *InfinityJ urnal Ilmiah Program Studi Matematika STKIP Siliwangi Bandung*, Vol. 3, Issue 2. *https://e-journal.stkipsiliwangi.ac.id/index.php/infinity/article/view/59.*

Riani, R. A., & Sofra, A. Y. (2023). Pengelompokan berdasarkan garis kemiskinan pendekatan time series based clustering di Provinsi Jawa Timur. *Mathunesa: Jurnal Ilmiah Matematika*, 11(3), 478–488.

Sachrrial, R. H., & Iskandar, A. (2023). Analisa Perbandingan Complate Linkage AHC dan K-Medoids Dalam Pengelompokkan Data Kemiskinan di Indonesia. *Building of Informatics, Technology and Science (BITS)*, *5*(2). https://doi.org/10.47065/bits.v5i2.4310

Salsabila, F., Ridwan, T., & H, H. (2024). Analisa Volume Penyebaran Sampah Di Karawang Menggunakan Algoritma K-Means Clustering. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(2). https://doi.org/10.23960/jitet.v12i2.4226

Septianingsih, A. (2022). Pemetaan Kabupaten Kota Di Provinsi Jawa Timur Berdasarkan Tingkat Kasus Penyakit Menggunakan Pendekatan Agglomeratif Hierarchical Clustering. *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, *3*(2), 367–386. https://doi.org/10.46306/lb.v3i2.139

Suhirman, S., & Wintolo, H. (2019). System for Determining Public Health Level Using the Agglomerative Hierarchical Clustering Method. *Compiler*, *8*(1), 95. https://doi.org/10.28989/compiler.v8i1.425

Suraya, G. R., & Wijayanto, A. W. (2022). Perbandingan metode hierarchical clustering, K-Means, K-Medoids, dan fuzzy C-Means dalam pengelompokan provinsi di Indonesia menurut indeks khusus penanganan stunting. *Indonesian Journal of Statistics and Its Applications*, 6(2), 180–201.

Syafiyah, U., Puspitasari, D. P., Asrafi, I., Wicaksono, B., & Sirait, F. M. (2022, November). Analisis perbandingan hierarchical dan non-hierarchical

clustering pada data indikator ketenagakerjaan di Jawa Barat tahun 2020. In *Seminar Nasional Official Statistics* (Vol. 2022, No. 1, pp. 803–812).

Tarigan, M. S. (2024). Pengelompokkan data rekam medis pada penyakit diabetes menggunakan metode divisive analysis clustering. *Sistemasi: Jurnal Sistem Informasi*, 13(4).

Tias, F. H., & Arum Handini Primandari. (2023). Pengelompokan Kecamatan di Kabupaten Merangin Berdasarkan Produksi Tanaman Perkebunan Tahun 2021 Menggunakan Agglomerative Hierarchical Clustering. *Emerging Statistics and Data Science Journal*, *1*(1), 137–147. https://doi.org/10.20885/esds.vol1.iss.1.art15

Ulinnuha, N., & Veriani, R. (2020). Analisis *Cluster* dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode *Complete Linkage, Average Linkage dan Ward*. *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, No.1, Vol.5, 102-108, : *https://core.ac.uk/download/pdf/328277185.pdf.*

Utami, N. D. (2017). Perbandingan hasil pengelompokkan antara metode average linkage, Ward, complete linkage, dan single linkage *(Studi kasus:* Indikator kesehatan Indonesia tahun 2015).

Wahyuni, S. (2021). Pengaruh pengangguran terbuka terdidik universitas terhadap garis kemiskinan di Provinsi Aceh. *Ekonomika: Jurnal Ekonomi dan Pembangunan*, 13(1), 8–12.

Widyadhana, D., Hastuti, R. B., Kharisudin, I., & Fauzi, F. (2021). Perbandingan Analisis Klaster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah. *PRISMA, Prosiding Seminar Nasional Matematika*, *4*(2), 584–594.