



## Klasterisasi Pola Gejala Depresi Menggunakan *Agglomerative Hierarchical Cluster Analysis* Berdasarkan Skor Depresi PHQ-9

Ermila<sup>1\*</sup>, Ghefira Zahra Nur Fadhilah<sup>2</sup>, Ni Wayan Erdiani<sup>3</sup>, Rizal Adi Saputra<sup>4</sup>  
<sup>1,2,3,4</sup> Universitas Halu Oleo, Indonesia

Kampus Hijau Bumi Tridharma, Anduonohu, Kec. Kambu,  
Kota Kendari, Sulawesi Tenggara

Email : [ermila304@gmail.com](mailto:ermila304@gmail.com) [ghefirazahranurfadhilah@gmail.com](mailto:ghefirazahranurfadhilah@gmail.com)  
[niwayanerdianizz@gmail.com](mailto:niwayanerdianizz@gmail.com) [rizaladisaputra@uho.ac.id](mailto:rizaladisaputra@uho.ac.id)

**Abstract.** *Mental health issues such as stress, anxiety, and depression are increasingly affecting individuals across various levels, from mild to severe. Despite this, public perception often underestimates these conditions, particularly depression, which is frequently misunderstood as a temporary emotional phase. In fact, depression is a serious mental disorder that requires proper attention and treatment. This study aims to classify early patterns of depressive symptoms using the Agglomerative Hierarchical Clustering (AHC) algorithm. Data were collected from 1,030 respondents using the PHQ-9 questionnaire, a widely recognized tool for assessing depression levels. The analysis was conducted using the AHC method, and cluster evaluation was performed using the Davies-Bouldin Index (DBI) to determine optimal clustering. Results indicate that the Complete Linkage method provided the most optimal performance, as reflected in the highest Cophenetic Correlation Coefficient compared to other linkage methods. The DBI evaluation suggests that the optimal number of clusters is four. Among the respondents, 17.2% fall into cluster 1, 32.5% into cluster 2, 21.9% into cluster 3, and 28.3% into cluster 4. These clusters represent distinct patterns of depressive symptoms, which can be used for early detection and improved mental health interventions.*

**Keywords:** *AHC, Clustering, Depression, Mental Health, PHQ-9*

**Abstrak.** Banyak individu saat ini menghadapi tingkat stres, kecemasan, dan kegelisahan mulai dari yang umum hingga sudah ditahap parah. Namun, persepsi umum bahwa stres dan depresi tidak merupakan gangguan mental masih banyak ditemui. Depresi sering dipandang sebagai kondisi yang ringan dan dianggap membaik dengan sendirinya, padahal kenyataannya depresi merupakan suatu gangguan yang lebih dari sekadar perubahan emosi. Kondisi ini bukanlah sesuatu yang bisa diselesaikan dengan cepat dan instan. Berdasarkan hal tersebut, penelitian ini bertujuan untuk mengelompokkan pola gejala awal depresi dengan menggunakan algoritma Agglomerative Hierarchical Clustering. Metodologi yang digunakan dari penelitian ini dimulai dari pengumpulan data yang di dapatkan dari kuesioner PHQ-9 dengan responden sebanyak 1030 data, kemudian cluster dengan metode AHC dan DBI untuk evaluasi. Hasil yang didapatkan dari penelitian ini menunjukkan bahwa algoritma yang paling optimal untuk penelitian ini adalah Complete Linked berdasarkan Nilai Cophenetic Correlation nya yang lebih tinggi dibanding metode lainnya. Berdasarkan hasil evaluasi DBI didapatkan pembentukan cluster optimal adalah 4 cluster. Dari total responden, 17,2% berada pada cluster 1, 32,5% berada pada cluster 2, 21,9% berada pada cluster 3, dan 28,3% berada pada cluster 4.

**Kata kunci:** AHC, Depresi, Kesehatan Mental, Klasterisasi, PHQ-9

### 1. LATAR BELAKANG

Depresi adalah suatu kondisi terganggunya mental yang ditandai dengan perasaan murung yang terus-menerus, keputusasaan, rasa bersalah, serta perasaan tidak berharga. Ketika stres dan depresi tidak ditangani, hal ini bisa berdampak negatif pada sistem kekebalan tubuh. Sistem kekebalan tubuh kita melemah ketika kita mengalami emosi yang tidak menyenangkan seperti merasa iri, murung, marah, putus asa, gelisah, dan kurang bersyukur. Orang yang mengalami depresi biasanya menampilkan gejala-gejala yang umum dalam aspek fisik, psikologis, dan sosial. Sementara beberapa orang menunjukkan sedikit gejala, yang lain menunjukkan banyak gejala. Seiring waktu, intensitas gejala-gejala tersebut berubah.

World Health Organization (WHO) menyebutkan bahwa depresi adalah salah satu penyakit mental yang biasanya ditandai oleh berbagai gejala seperti suasana hati yang depresif, hilangnya minat atau kesenangan, kurangnya energi, perasaan bersalah atau rendah diri, kesulitan tidur, berkurangnya nafsu makan, kelelahan, dan masalah dalam berkonsentrasi. Kondisi ini bisa semakin memperburuk seiring berjalannya waktu dan mengalami fase kambuh yang berulang, sehingga dapat membatasi kemampuan seseorang untuk melakukan kegiatan sehari-hari. Organisasi Kesehatan Dunia (WHO) juga memperkirakan bahwa di seluruh dunia, terdapat 280 juta orang mengalami depresi, dan hal ini menjadi penyebab utama kematian akibat bunuh diri di kalangan individu usia 15 hingga 29 tahun.

Menurut data statistik dari Riset Kesehatan Dasar (Riskesdas) 2018, Indonesia menduduki peringkat kelima dengan angka kejadian gejala depresi yang tinggi, yaitu 21,8% pada masyarakat Indonesia, dengan 6,1% orang berusia 15 tahun ke atas melaporkan adanya gejala depresi dan hanya 9% yang mendapatkan pengobatan[4]. Sedangkan menurut data dari Indonesia National Adolescent Mental Health Survey (I-NAMHS) yang melakukan survey pada tahun 2022, masalah kesehatan mental mempengaruhi satu dari tiga remaja di Indonesia dengan depresi mayor memiliki prevalensi tertinggi kedua dari masalah ini, yaitu sebesar 1,0% dari total sampel. Ikatan Dokter Anak Indonesia cabang Yogyakarta juga mengungkapkan sekitar 3% dari anak-anak usia sekolah dan 6% dari remaja menderita penyakit depresi yang serius. Anak-anak dan remaja yang mengalami depresi menunjukkan gangguan yang signifikan dalam kinerja mereka di keluarga, sekolah, dan dalam bergaul dengan teman-teman mereka.

Terdapat hasil penelitian yang mengumpulkan data tweet menggunakan kata kunci depresi berbasis phq-9 yang mendapatkan 2581 data, di mana 924 di antaranya diklasifikasikan sebagai depresi ringan, 397 sebagai depresi sedang, dan 1260 sebagai depresi berat. Terdapat penelitian lain terkait depresi pada mahasiswa dengan 60 responden yang diukur, 40% tidak menunjukkan tanda-tanda depresi, 45% menunjukkan depresi sedang, dan 15% menunjukkan depresi berat. Penelitian lain yang menggunakan data dari total 24 pasien hepatitis C yang mengisi kuesioner PHQ-9. Hasil penelitian menunjukkan bahwa terdapat 9 pasien dengan depresi minimal (37,5%), 12 pasien dengan depresi ringan (50%), 2 pasien dengan depresi sedang (8,3%), dan 1 pasien dengan depresi sedang-berat (4,2%).

Salah satu pendekatan dalam data mining untuk mengelompokkan data adalah menggunakan metode clustering. Proses pengorganisasian objek ke dalam kelompok-kelompok sehingga objek-objek tertentu dalam satu klaster memiliki karakteristik yang sama dan objek-objek lainnya berbeda dengan objek-objek di klaster lain dikenal sebagai clustering. Clustering sering dianggap sebagai tugas pembelajaran tanpa pengawasan karena dalam

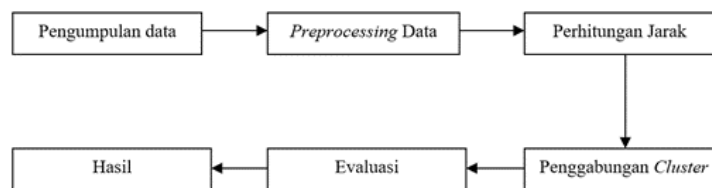
prosesnya, tidak terdapat label atau kelas yang sudah ditentukan sebelumnya untuk sampel data. Dalam clustering, ada dua pendekatan yang berbeda yaitu clustering partisi dan clustering hirarki. Salah satu manfaat dari pengelompokan hierarkis adalah dapat menggabungkan data dan menghasilkan hierarki data yang mirip akan disusun dalam hierarki terdekat, dan data yang tidak terkait akan disusun dalam hierarki yang jauh. Metode pengelompokan hirarki ini melalui proses pembuatan basis pohon biner, yang juga dikenal sebagai dendrogram. Agglomerative dan Divisive adalah dua pendekatan dari Hierarchical Clustering. Pendekatan Agglomerative membangun hirarki cluster dari bawah ke atas dengan menggabungkan dua cluster pada satu waktu, dimulai dengan satu cluster (yang hanya menyimpan satu objek data per cluster) di tingkat bawah. Terdapat lima algoritma dalam Agglomerative, yaitu Single Linkage, Complete Linkage, Average Linkage, Centroid Linkage dan Ward's Linkage.

Terdapat penelitian mengenai hasil pengujian dan analisis sistem kesehatan masyarakat dengan menggunakan pendekatan Agglomerative Hierarchical Clustering (AHC). Sistem yang telah dikembangkan dapat secara efektif menyelesaikan kerumitan dalam menentukan jarak terjauh dengan memanfaatkan metode Agglomerative Hierarchical Clustering (AHC) dengan metode penggabungan cluster menggunakan Complete Linkage (jarak terjauh). Selain itu, penelitian lain telah menghasilkan hasil klaster yang memberikan informasi yang berguna untuk membedakan ciri-ciri individu dengan gejala depresi dan asma. Variasi yang signifikan dalam skor ASK-12 antara Klaster A, B, dan C telah ditemukan melalui pengelompokan. Dengan demikian, dengan menggunakan fitur pengobatan pasien, hasil pengelompokan ini dapat membantu pengelompokan pasien.

Tingkat kejadian depresi yang semakin meningkat menunjukkan perlunya penanganan serius agar dapat dikurangi secara efektif. Depresi harus segera diatasi karena gejala depresi berdampak pada perkembangan sumber daya manusia di Indonesia, sehingga pencegahan dan penanganan gejala depresi harus diperhatikan sejak dini bahkan sejak munculnya tanda-tanda awal depresi. Untuk mengatasi masalah kesehatan mental di Indonesia, penelitian ini melakukan clusterisasi pada pola gejala depresi menggunakan Agglomerative Hierarchical Cluster berdasarkan skor depresi PHQ-9. Secara umum, output penelitian diharapkan dapat menghasilkan pola cluster yang dapat digunakan untuk menggambarkan bagaimana data dikelompokkan berdasarkan gejala yang muncul. Clusterisasi pola gejala depresi menggunakan Hierarchical Cluster Analysis berdasarkan skor depresi PHQ-9 ini akan memberikan intervensi yang lebih efektif untuk penanganan dan pencegahan kasus depresi.

## 2. METODE

Dalam penelitian ini proses pembuatan modelnya menggunakan metode Agglomerative Hierarchical Cluster yang alur prosesnya dimulai dari pengumpulan data hingga tahap hasil. Berikut adalah tahapan dari metode Agglomerative Hierarchical Cluster pada Gambar 1.



Gambar 1. Tahapan metode *Agglomerative Hierarchical Cluster*

### Pengumpulan Data

Metode yang digunakan dalam penelitian ini adalah pengumpulan data melalui kuesioner daring yang terdiri dari 12 pertanyaan. Kuesioner ini dibagi menjadi dua bagian utama, yaitu pertanyaan data diri dan pertanyaan terkait PHQ-9 (Patient Health Questionnaire-9). Bagian pertama dari kuesioner mencakup 3 pertanyaan yang bertujuan untuk mengumpulkan informasi demografis dasar seperti nama/inisial, usia dan jenis kelamin. Bagian kedua dari kuesioner terdiri dari 9 pertanyaan yang merupakan bagian dari PHQ-9. PHQ-9 merupakan serangkaian pertanyaan yang digunakan untuk menentukan tingkat keparahan awal depresi selama proses skrining awal. PHQ-9 adalah alat penilaian yang didesain untuk mengukur intensitas gejala depresi dan menyaring kemungkinan adanya gangguan depresi. Pertanyaan-pertanyaan dalam PHQ-9 dirancang untuk mengevaluasi gejala yang dialami individu selama dua minggu terakhir, seperti suasana hati yang sedih, kehilangan minat atau kesenangan dalam melakukan aktivitas, perubahan pola tidur dan nafsu makan, perasaan lelah atau kurang energi, dan kesulitan berkonsentrasi. Skor dari PHQ-9 memberikan indikasi mengenai sejauh mana gejala-gejala tersebut mempengaruhi keseharian dan kualitas hidup individu, serta dapat membantu dalam merencanakan intervensi yang tepat[12]. pertanyaan kuesioner PHQ-9 dapat dilihat pada Tabel 1.

Tabel 1. Pertanyaan Kuesioner PHQ-9

No	Dalam dua minggu terakhir, seberapa sering kamu merasa terganggu oleh hal-hal beriku	Tidak Pernah	Beberapa Hari	Lebih dari Separuh Waktu	Hampir Setiap Hari
1.	Kurangnya minat atau tertarik dalam kegiatan sehari-hari?	0	1	2	3
2.	Merasa sedih, tertekan, atau putus asa?	0	1	2	3
3.	Sulit tidur atau tidur berlebihan?	0	1	2	3

4.	Merasa lelah atau kekurangan energi?	0	1	2	3
5.	Nafsu makan buruk atau makan berlebihan?	0	1	2	3
6.	Sering merasa tidak berharga, merasa gagal, atau merasa mengecewakan diri sendiri atau keluarga?	0	1	2	3
7.	Sulit berkonsentrasi?	0	1	2	3
8.	Lambat bergerak atau berbicara, sehingga orang lain memperhatikannya? kurang istirahat dan sulit diam lebih dari biasanya?	0	1	2	3
9.	Berpikir bahwa lebih baik mati, memiliki pikiran untuk menyakiti diri sendiri dengan cara tertentu?	0	1	2	3

Tingkat depresi yang diukur melalui total skor jawaban dari seluruh pertanyaan dalam kuesioner PHQ-9 digunakan untuk mengklasifikasikan tingkat keparahan depresi ke dalam empat kategori berbeda. Tingkat keparahan ini terbagi menjadi empat kategori yang spesifik, yaitu:

- Total skor 0-4 termasuk dalam kategori normal/tidak ada depresi.
- Total skor 5-9 termasuk dalam kategori gejala depresi ringan.
- Total skor 10-19 termasuk dalam kategori gejala depresi sedang.
- Total skor 20-27 termasuk dalam kategori gejala kecemasan Berat.

### **Preprocessing Data**

*Preprocessing* adalah langkah awal dalam pemrosesan data yang melibatkan pembersihan, transformasi, dan normalisasi data untuk memastikan kualitas dan konsistensi sebelum analisis lebih lanjut. Data mentah dari kuesioner PHQ-9 diproses melalui beberapa langkah untuk memastikan kualitas dan konsistensi data. Langkah pertama adalah pembersihan data, di mana *entri* yang tidak lengkap atau tidak valid dihapus untuk mencegah redudansi dan inkonsistensi. Selain itu, variabel yang tidak relevan atau tidak dapat dihitung juga dihilangkan, sehingga hanya data angka yang dapat diolah oleh metode *clustering* yang digunakan. Langkah berikutnya adalah normalisasi data. Meskipun skala PHQ-9 sudah seragam, normalisasi tetap dilakukan untuk memastikan setiap fitur (pertanyaan) memiliki skala yang sama. Penting untuk

memastikan hasil klasterisasi tetap konsisten dan akurat, karena metode *clustering* hanya bisa memproses data dalam bentuk angka.

### Perhitungan Jarak

*Agglomerative Hierarchical Clustering* (AHC) adalah sebuah metode data mining untuk pembentukan kelompok (*clustering*). *Agglomerative hierarchical clustering* (AHC) menggunakan pendekatan *bottom-up*, setiap data dimulai sebagai klaster terpisah. Kemudian, secara rekursif, *cluster-cluster* ini mencari pasangan terdekat untuk digabungkan menjadi *cluster* yang lebih besar. Dalam metode AHC, Proses pembentukan *cluster* diawali dengan menganggap setiap objek sebagai *cluster* yang terpisah atau *cluster* atom. Kemudian, *cluster-cluster* tersebut digabungkan satu sama lain untuk membentuk *cluster* yang lebih besar. Proses penggabungan dalam *clusterisasi* hierarkis ini berlangsung secara iteratif, di mana pada setiap tahap, dua *cluster* yang paling mirip (atau paling dekat dalam hal jarak) digabungkan menjadi satu *cluster* yang lebih besar. Proses ini berlanjut hingga semua objek atau *cluster* terhubung menjadi satu *cluster* tunggal, atau sampai kondisi batas tertentu yang telah ditentukan sebelumnya tercapai dan proses penggabungan berakhir. Pada setiap tahap penggabungan, dua objek atau *cluster* (disebut sebagai *atomic clusters*) digabungkan berdasarkan ukuran jarak yang dikenal sebagai *Euclidean Distance*. Jarak *Euclidean* merupakan ukuran jarak lurus antara dua titik dalam ruang dimensi yang diamati. Terdapat objek pertama yang diamati adalah  $X = (x_1, x_2, x_3, \dots, x_p)$  dan  $Y = (y_1, y_2, y_3, \dots, y_p)$ . Untuk menghitung jarak antara objek-objek yang akan digabungkan, digunakan persamaan berikut:

$$d_{(x,y)} = \sqrt{\sum_{j=1}^p (x_j - y_j)^2} \quad (1)$$

Keterangan:

$d_{(x,y)}$  : Jarak *euclidean* antara objek  $X$  dengan objek  $Y$

$P$  : Banyaknya variabel yang diamati

$x_j$  : Nilai  $j$  pada objek  $X$

$y_j$  : Nilai  $j$  pada objek  $Y$

### Penggabungan Cluster

Dalam analisis *cluster*, penggabungan *cluster* dengan *agglomerative clustering* adalah metode hierarkis yang mengelompokkan data dengan melakukan penggabungan berulang dari *cluster-cluster* kecil menjadi *cluster* yang lebih besar hingga semua data tergabung dalam satu *cluster*. Untuk menentukan metode penggabungan *cluster* terbaik bisa menggunakan *correlation cophenetic*. *Correlation cophenetic* adalah ukuran yang digunakan untuk mengevaluasi seberapa baik dendrogram yang dihasilkan oleh analisis *cluster* hierarkis

mencerminkan data aslinya. *Correlation cophenetic* adalah nilai korelasi yang mengukur hubungan antara elemen-elemen dalam matriks ketidakmiripan asli (yang didasarkan pada jarak *Euclidean*) dan elemen-elemen dalam matriks *cophenetic* yang dihasilkan. *Correlation cophenetic* memiliki rentang nilai dari -1 hingga 1, dengan nilai yang mendekati 1 menunjukkan bahwa dendrogram secara akurat dapat merepresentasikan jarak antara pasangan data asli. Berikut adalah persamaan *cophenetic correlation*.

$$r_{coph} = \frac{\sum_{i < k} (d_{ik} - \bar{d})(d_{Cik} - \bar{d}_c)}{\sqrt{[\sum_{i < k} (d_{ik} - \bar{d})^2][\sum_{i < k} (d_{Cik} - \bar{d}_c)^2]}} \quad (2)$$

Keterangan:

$r_{coph}$  : Koefisien korelasi *cophenetic*

$d_{ik}$  : Jarak *Euclidean* objek ke-i dan ke-k

$\bar{d}$  : Rata-rata  $d_{ik}$

$d_{Cik}$  : Jarak *Cophenetic* objek ke-i dan ke-k

$\bar{d}_c$  : Rata-rata  $d_{Cik}$

Metode *Agglomerative clustering* digunakan untuk mengelompokkan data dengan cara menggabungkan entitas-entitas yang mirip secara bertahap menjadi kelompok yang lebih besar. Proses ini dimulai dengan setiap titik dianggap sebagai cluster individual, kemudian secara berurutan menggabungkan dua *cluster* yang paling dekat (atau paling mirip) berdasarkan suatu metrik jarak yang telah ditentukan. Terdapat lima metode *agglomerative* yang dapat digunakan untuk membentuk kluster, yaitu *single Linkage*, *Average Linkage*, *Complete Linkage*, *Centroid Linkage* dan *ward's Linkage*. Metode *agglomerative clustering* dapat diterapkan untuk membentuk *cluster* berdasarkan dengan proses berikut.

- *Single Linkage Method* atau yang juga dikenal sebagai *single link* merupakan salah satu teknik penggabungan kluster dalam *Agglomerative Hierarchical Clustering* (AHC) di mana objek-objek dikelompokkan berdasarkan jarak terdekat antara satu objek dengan yang lain. Dengan kata lain, dua *cluster* akan digabungkan jika terdapat dua titik, masing-masing dari *cluster* yang berbeda, yang memiliki jarak terpendek di antara semua pasangan titik yang mungkin. Untuk memulai pengelompokan, pertama-tama kita harus menentukan nilai minimum dari  $D = \{d_{(xy)}\}$  dan kemudian menggabungkan objek-objek yang berdekatan satu sama lain, seperti  $U$  dan  $V$  dan untuk membuat sebuah *cluster* ( $UV$ ). Lalu selanjutnya pemisahan antara *cluster* dan objek lain menggunakan rumus berikut.

$$d_{(uv)w} = \min(d_{uw}, d_{vw}) \quad (3)$$

- *Complete Linkage Method* (*Complete link*) adalah salah satu teknik penggabungan kluster AHC yang membangun cluster berdasarkan jarak maksimum antara setiap pasangan objek.

Dalam *Complete Linkage*, jarak antara dua *cluster* dihitung sebagai jarak terjauh antara setiap pasangan objek yang berasal dari dua *cluster* yang berbeda. Langkah awal dalam pengelompokan adalah mencari nilai terkecil dari  $D = \{d_{(xy)}\}$  dan mengelompokkan kluster-kluster yang berdekatan satu sama lain. Misalnya,  $U$  dan  $V$  yang berdekatan satu sama lain dan membentuk sebuah *cluster* ( $UV$ ). Selanjutnya itu, pada langkah pemisahan antara *cluster* dengan objek lain menggunakan rumus berikut.

$$d_{(uv)w} = \max(d_{uw}, d_{vw}) \quad (4)$$

- *Average Linkage Method (Average link)* adalah teknik penggabungan objek pada AHC yang dapat digunakan untuk mengelompokkan objek berdasarkan rata-rata jarak antara satu objek dengan yang lain. Teknik ini menghitung jarak antara dua *cluster* dihitung dengan mengambil nilai rata-rata dari jarak antara setiap pasangan objek yang terdiri dari satu objek di *cluster* pertama dan satu objek di *cluster* kedua. Untuk membuat *cluster*, pertama-tama tentukan nilai minimum dari  $D = \{d_{(xy)}\}$  dan kemudian gabungkan objek-objek yang berdekatan satu sama lain, seperti  $U$  dan  $V$  yang berdekatan satu sama lain dan membentuk sebuah *cluster* ( $UV$ ). Selanjutnya itu, pada langkah pemisahan jarak antara *cluster* dengan objek lain menggunakan rumus berikut.

$$d_{(uv)w} = \text{Average}(d_{uw}, d_{vw}) \quad (5)$$

- *Centroid Linkage Method* adalah salah satu teknik penggabungan objek pada AHC yang menggabungkan *cluster* berdasarkan jarak antara *Centroid* (titik pusat) dari *cluster* yang berbeda. Dalam metode ini, setiap kali dua *cluster* digabungkan, *Centroid* baru dihitung sebagai rata-rata dari semua titik dalam *cluster* baru yang terbentuk. Metode ini memperhitungkan ulang *centroid* setiap kali terbentuk *cluster* baru, sehingga *cluster-cluster* dapat stabil terbentuk. Jarak antara setiap kelompok ( $UV$ ) dengan  $W$  adalah.

$$d_{(uv)w} = \text{median}(d_{uw}, d_{vw}) \quad (6)$$

- *Ward's Linkage Method* adalah metode AHC yang membentuk *cluster* dengan cara meminimalkan jumlah variansi dalam setiap *cluster*. Setiap kali dua *cluster* digabungkan, *Ward's Method* memilih pasangan *cluster* yang menghasilkan peningkatan terkecil dalam jumlah kuadrat kesalahan (*squared error*) dalam *cluster*. Metode ini berusaha mengurangi variasi di dalam setiap kelompok dan sering digunakan untuk menggabungkan kelompok-kelompok dengan jumlah yang sedikit.

$$ESS = \sum_{i=1}^N (x_j - \bar{x}) (x_j - \bar{x}) \quad (7)$$

## Evaluasi

Selanjutnya adalah tahap evaluasi yang dalam *clusterisasi* digunakan untuk memastikan bahwa *cluster* yang dihasilkan valid. Dalam penelitian ini metode evaluasi yang



digunakan adalah Davies-Bouldin Index (DBI). DBI menghitung perbandingan antara jarak di dalam kluster dengan jarak antar kluster. DBI adalah sebuah metrik evaluasi yang digunakan untuk menilai seberapa baik kualitas dari *cluster* yang terbentuk. berdasarkan jarak antara *cluster*, dan semakin rendah nilainya menunjukkan semakin baik *clusterisasi* tersebut. DBI dihitung dengan cara membandingkan jarak antara pusat *cluster* (*Centroid*) dengan *cluster* lainnya dalam konteks ruang fitur.

Nilai DBI yang lebih rendah menunjukkan bahwa *cluster-cluster*nya lebih terpisah dan lebih kompak. Setelah melakukan *clusterisasi* menggunakan dendrogram dan metode *Complete Linkage*, kita akan menghitung DBI untuk berbagai jumlah *cluster* (biasanya dari 2 hingga 10 *cluster*) dan memilih jumlah *cluster* yang memberikan nilai DBI terendah, karena ini menandakan *clusterisasi* yang lebih baik. Berikut ini adalah langkah-langkah untuk menghitung DBI.

- Menghitung jarak *intra-cluster*, standar deviasi dihitung antara setiap titik data  $X_j$  dalam satu *cluster*  $i$  terhadap pusat *cluster*  $A_i$ , jumlah data dalam *cluster*  $i$  direpresentasikan oleh  $T_i$ ,

$$S_i = \left\{ \frac{1}{T_i} \sum_{j=1}^{T_i} |X_j - A_i|^2 \right\}, \quad (8)$$

- Menghitung jarak antara setiap pasang *cluster* dengan menggunakan pusat  $a_i$  dari *cluster*  $i$  dan  $a_j$  dari *cluster*  $j$ , dimana  $i \neq j$ ,

$$M_{ij} = \|a_i - a_j\|^2, \quad (9)$$

- Untuk setiap pasangan *cluster*, sebaran dari masing-masing *cluster* dijumlahkan, kemudian hasilnya dinormalisasi dengan membagi totalnya dengan jarak antara kedua *cluster*. Metode ini memberikan ukuran seberapa baik kedua *cluster* tersebut dikelompokkan bersama,

$$R_{ij} = \frac{S_i + S_j}{M_{ij}}, \quad (10)$$

- Menentukan nilai maksimum dari ukuran kesamaan antar kelompok,

$$R_i = \max_{i \neq j} R_{ij} \quad (11)$$

- Menghitung DBI sebagai rata-rata dari tingkat kesamaan maksimum *cluster*, dengan  $K$  adalah banyaknya *cluster*.

$$DBI = \frac{1}{k} \sum_{i=1}^k R_i \quad (12)$$

## Hasil

*Dendrogram* adalah gambaran grafis dari prosedur sebuah analisis *clustering* yang menunjukkan bagaimana *cluster* terbentuk dan nilai koefisien jaraknya di setiap tahap. Objek-objeknya adalah angka-angka di sebelah kanan. Objek-objek tersebut merupakan objek-objek

penelitian, dimana objek-objek tersebut tersusun dalam sebuah cluster dengan objek-objek lainnya melalui hubungan yang dibuat oleh garis-garis. Objek-objek yang akan bergabung membentuk satu *cluster* ditunjukkan oleh percabangan pada dendrogram[16]. Dalam dendrogram, setiap cabang mewakili penggabungan dua *cluster* berdasarkan nilai koefisien jarak (atau kesamaan) yang dihitung berdasarkan metode *agglomerative clustering* yang dipilih. Pada setiap tahap penggabungan, nilai koefisien jarak menunjukkan seberapa jauh atau seberapa mirip dua *cluster* tersebut sebelum digabungkan menjadi satu *cluster* baru. Proses ini dilanjutkan hingga semua objek tergabung dalam satu *cluster* utama yang besar atau jumlah *cluster* yang diinginkan telah terbentuk.

### **3. HASIL DAN PEMBAHASAN**

Penelitian ini bertujuan untuk mengelompokkan pola gejala depresi berdasarkan skor PHQ-9, yang merupakan alat penilaian standar untuk mengukur tingkat keparahan depresi. Proses pengelompokan dilakukan dengan metode Agglomerative Hierarchical Clustering (AHC) dalam data mining, yang memungkinkan identifikasi struktur alami dalam data dan pengelompokan individu dengan karakteristik gejala yang serupa. Data dari sembilan kriteria yang diukur oleh PHQ-9 diolah dan dianalisis untuk membentuk cluster yang berbeda. Data dari sembilan kriteria yang diukur oleh PHQ-9 diolah dan dianalisis untuk membentuk cluster yang berbeda dengan tujuan untuk memahami gejala depresi dan mengidentifikasi pengelompokan data.

#### **Pengumpulan Data**

Dataset yang telah dikumpulkan melalui distribusi kuesioner daring, menghasilkan data dari 1.030 responden yang berpartisipasi. Setiap entri dalam dataset ini mencatat informasi dari 12 atribut yang berbeda, membentuk total 1.030 record. Data yang telah terverifikasi sebagai bersih dan lengkap, tanpa adanya kekosongan nilai (*missing value*)s selanjutnya langsung masuk ke tahap selanjutnya yaitu data preprocessing.

#### **Preprocessing Data**

Pada tahap ini dilakukan penyaringan atribut dengan mempertahankan hanya 9 pertanyaan PHQ-9, sementara atribut lainnya dihapus. Atribut yang dianggap tidak relevan atau tidak diperlukan untuk analisis, seperti Timestamp, Nama, Usia, dan Jenis Kelamin dihapus/dihilangkan untuk menyederhanakan dataset dan fokus hanya pada informasi yang esensial. Dengan demikian, dataset saat ini terdiri dari 1030 entri data dan 9 atribut yang relevan. Pada tahap ini, juga dilakukan perubahan nama kolom untuk pertanyaan PHQ-9 menjadi Q1 hingga Q9 untuk menyederhanakan atau mempersingkat nama kolom.

Tabel 2. Hasil Preprocessing Data

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
1	1	1	1	1	1	1	1	0	1
2	1	2	2	1	2	2	2	2	1
3	2	3	2	0	0	3	2	0	3
4	0	0	1	0	3	3	3	3	3
5	1	1	1	2	2	1	1	1	1
...	...	...	...	...	...	...	...	...	...
6	1	2	3	2	1	2	2	1	0
7	3	3	2	3	3	3	1	3	0
8	2	2	2	3	1	2	2	1	1
9	2	3	3	3	3	3	3	2	2
10	3	3	3	3	3	3	3	3	3

### Perhitungan Jarak

Setelah menyelesaikan tahap preprocessing data, langkah berikutnya adalah menggunakan metode Euclidean Distance untuk mengukur jarak antara setiap data. Metode ini menghitung jarak garis lurus (straight-line distance) antara dua titik dalam ruang multidimensi, yang ditentukan oleh atribut-atribut yang dimiliki data. Hasil perhitungan ini akan memberikan gambaran tentang seberapa jauh atau dekatnya setiap data dalam ruang tersebut, memungkinkan kita untuk mengidentifikasi pola, kesamaan, dan perbedaan di antara data. Berikut adalah matriks jarak Euclidean distance yang dihasilkan dari proses ini pada Tabel 3.

Tabel 3. Matriks Jarak *Euclidean*

	Q1	Q2	Q3	Q4	Q5	...	Q1026	Q1027	Q1028	Q1029	Q1030
Q1	0.00	3.00	4.12	5.29	1.73	...	3.16	5.56	3.16	5.47	6.40
Q2	3.00	0.00	4.00	3.87	2.44	...	2.23	3.74	2.64	3.32	4.24
Q3	4.12	4.00	0.00	5.74	4.89	...	4.35	6.16	4.12	5.00	5.47
Q4	5.29	3.87	5.74	0.00	4.79	...	5.65	6.40	5.65	5.29	5.56
Q5	1.73	2.44	4.89	4.79	0.00	...	3.00	4.47	2.64	4.58	5.47
...	...	...	...	...	...	...		...	...	...	...
Q1026	3.16	2.23	4.35	5.65	3.00		0.00	4.12	2.00	3.74	5.00
Q1027	5.56	3.74	6.16	6.40	4.47	...	4.12	0.00	3.60	3.31	3.74
Q1028	3.16	2.64	1.12	5.65	2.64	...	2.00	3.60	0.00	3.16	4.12
Q1029	5.47	3.31	5.00	5.29	4.58	...	3.74	3.31	3.16	0.00	1.73
Q1030	6.40	4.24	5.47	5.56	5.47	...	5.00	3.74	4.12	1.73	0.00

Berdasarkan Tabel 2, pasangan yang memiliki jarak *Euclidean* terkecil untuk membentuk cluster awal adalah Q1 dan Q48 dengan nilai 1.00. *Cluster* yang terbentuk dari pasangan Q1Q48 ini dinyatakan sebagai *cluster* dengan tingkat gejala depresi yang paling

rendah atau dinyatakan sebagai kelompok individu dengan kondisi kesehatan mental yang lebih baik.

### Penggabungan Cluster

Langkah berikutnya adalah mengelompokkan *cluster* berdasarkan jarak *Euclidean*. Proses pembentukan *cluster* menggunakan algoritma yang menghasilkan korelasi *cophenetic* tertinggi di antara algoritma *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Ward Linkage*, dan *Centroid Linkage*. Hasil dari beberapa algoritma pengelompokan tersebut ditampilkan dalam Tabel 4.

Tabel 4. *Chopenetic Correlation*

Metode	<i>Chopenetic Correlation</i>
<i>Single Linkage</i>	0.350
<i>Complete Linkage</i>	0.726
<i>Average Linkage</i>	0.670
<i>Ward Linkage</i>	0.603
<i>Centroid Linkage</i>	0.631

Pada tabel diatas menunjukkan bahwa metode yang memiliki nilai *correlation chopenetic* tertinggi adalah metode *Compelete Linkage* yaitu 0.726. Oleh karena itu, metode *Complete Linkage* dipilih untuk penelitian ini, karena tingginya korelasi *cophenetic* menunjukkan bahwa hasil *clustering* lebih representatif dan memiliki kesesuaian yang baik dengan jarak asli antar data. Selanjutnya, dalam klaster yang sudah ditetapkan sebagai satu kesatuan klaster yaitu Q1Q48 dilakukan pengelompokan atau penggabungan *cluster* selanjutnya dengan menggunakan metode *Complete Linkage*, di mana jarak antara satu *cluster* dengan *cluster* selanjutnya berdasarkan jarak terjauh antara setiap objek satu dengan yang lain. Berikut adalah matriks pembentukan *cluster* pertama pada Tabel 5.

Table 5. Matriks Jarak *Euclidean*

Kode	Q1Q48	Q2	Q3	Q4	Q5	...	Q1026	Q1027	Q1028	Q1029	Q1030
Q1Q48	0.00	3.00	4.12	5.29	1.73	...	3.16	5.56	3.16	5.47	6.40
Q2	3.08	0.00	4.00	3.87	2.44	...	2.23	3.74	2.64	3.32	4.24
Q3	4.41	4.00	0.00	5.74	4.89	...	4.35	6.16	4.12	5.00	5.47
Q4	5.52	3.87	5.74	0.00	4.79	...	5.65	6.40	5.65	5.29	5.56
Q5	1.87	2.44	4.89	4.79	0.00	...	3.00	4.47	2.64	4.58	5.47
...	...	...	...	...	...	...		...	...	...	...
Q1026	3.08	2.23	4.35	5.65	3.00		0.00	4.12	2.00	3.74	5.00
Q1027	5.52	3.74	6.16	6.40	4.47	...	4.12	0.00	3.60	3.31	3.74
Q1028	3.23	2.64	1.12	5.65	2.64	...	2.00	3.60	0.00	3.16	4.12
Q1029	5.61	3.31	5.00	5.29	4.58	...	3.74	3.31	3.16	0.00	1.73
Q1030	6.59	4.24	5.47	5.56	5.47	...	5.00	3.74	4.12	1.73	0.00

Dari analisis Tabel 5, terbentuk sebuah *cluster* baru dengan cara menghitung rata-rata jarak antar objek yang selanjutnya melibatkan pencarian nilai-nilai dalam matriks pada Tabel 5 untuk menentukan *cluster* berikutnya yang terhubung dengan *cluster* baru tersebut. Pendekatan ini memungkinkan untuk melacak hubungan antar *cluster* dan melakukan pembentukan *cluster* berdasarkan data yang ada. Langkah-langkah 3 dan 4 diulang hingga terbentuk satu *cluster* tunggal. Tahap selanjutnya ini adalah tahap evaluasi dengan DBI, untuk menentukan jumlah *cluster* optimal.

### Evaluasi

Pada tahap ini, dilakukan evaluasi dengan memanfaatkan *Davies-Bouldin Index* (DBI) untuk menentukan jumlah *cluster* yang optimal dari data yang telah di-*cluster*. DBI digunakan untuk mengukur kualitas partisi *cluster* berdasarkan jarak antar *cluster* dan dalam-*cluster*. Nilai DBI yang lebih rendah menandakan partisi yang lebih optimal, di mana *cluster-cluster* lebih padat dan terpisah dengan jelas di antara satu sama lain. Berikut adalah hasil perhitungan *Davies-Bouldin Index* (DBI) untuk 2 *cluster* hingga 10 *cluster* pada Tabel 6.

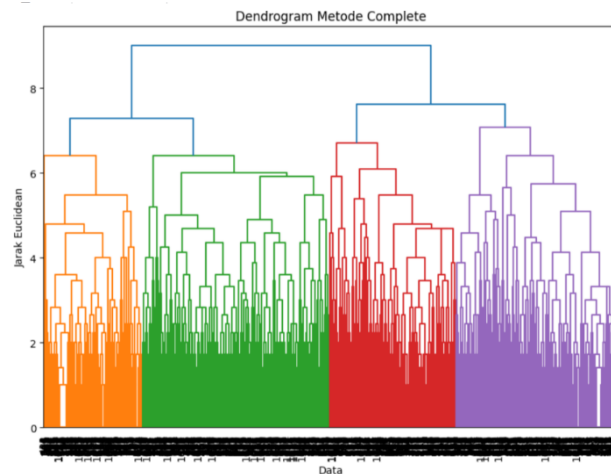
Tabel 6. Hasil DBI *cluster*

Jumlah <i>Cluster</i>	2	3	4	5	6	7	8	9	10
Hasil DBI	1.198	1.231	1.171	1.187	1.178	1.253	1.403	1.296	1.344

Berdasarkan tabel diatas, disimpulkan bahwa pembagian optimal adalah 4 *cluster* karena memiliki nilai DBI yang paling rendah, yaitu 1.171. DBI digunakan sebagai metrik evaluasi untuk menilai kualitas clustering, di mana semakin rendah nilainya menunjukkan bahwa *cluster-cluster* tersebut lebih padat dan terpisah dengan baik. Oleh karena itu, Ini menunjukkan bahwa partisi menjadi 4 *cluster* memberikan hasil yang lebih baik dalam hal kedekatan antar titik data dalam *cluster* dan pemisahan antar *cluster* dibandingkan dengan jumlah *cluster* lainnya yang dievaluasi.

### Hasil

Selanjutnya dalam analisis ini, dilakukan visualisasi dengan dendogram untuk menggambarkan hubungan hierarkis antara data. Dendogram ini memungkinkan untuk melihat struktur dan pola grup yang terbentuk berdasarkan metode pengelompokan yang digunakan. Berikut adalah hasil visualisasi dalam dendogram dari hasil pengelompokan pola gejala depresi berdasarkan skor PHQ-9 pada Gambar 2 berikut.



Gambar 2. Hasil Dendrogram

Hasil perhitungan *clustering* yang dilakukan menggunakan algoritma *Agglomerative Hierarchical Clustering* (AHC) dengan metode *Complete Linkage* menghasilkan empat *cluster* yang berbeda. Masing-masing *cluster* memiliki karakteristik unik yang mencerminkan pola dan struktur data yang berbeda. Berikut adalah jumlah anggota dalam setiap *cluster* tercantum dalam Tabel 7.

Tabel 7. Hasil Anggota *Cluster*

No	<i>Cluster</i>	Jumlah	Persentase
1.	<i>Cluster 1</i>	177	17.2%
2.	<i>Cluster 2</i>	335	32.5%
3.	<i>Cluster 3</i>	226	21.9%
4.	<i>Cluster 4</i>	292	28.3%

Dari hasil penelitian ini, dapat dirumuskan interpretasi komprehensif terhadap data yang dianalisis secara keseluruhan yaitu sebagai berikut.

- *Cluster 1*, ini terdiri dari 177 anggota yang mencakup individu-individu dengan gejala depresi yang lebih ringan dan bisa dibilang normal dibandingkan dengan *cluster* lainnya.
- *Cluster 2*, jumlah anggota *cluster* kedua adalah yang terbesar, dengan 335 anggota. Besarnya *cluster* ini menunjukkan bahwa ada banyak individu yang memiliki pola gejala yang serupa, yang menunjukkan gejala depresi ringan. Ini bisa menjadi *cluster* yang paling *heterogen* dengan variasi gejala yang luas.
- *Cluster 3*, ini terdiri dari 226 anggota. Anggota dalam *cluster* ini menunjukkan gejala depresi yang lebih sedang. *Cluster* ini dapat mencerminkan kelompok individu yang memiliki gejala yang lebih intens daripada *cluster* pertama dan kedua.
- *Cluster 4*, ini terdiri dari 292 anggota. Anggota dalam *cluster* ini mungkin menunjukkan gejala yang lebih parah atau spesifik yang berbeda dari *cluster* lainnya.

#### 4. KESIMPULAN

Depresi adalah penyakit mental yang ditandai dengan gejala sederhana seperti suasana hati tertekan, kesulitan tidur, penurunan nafsu makan, kelelahan, dan kesulitan fokus. Namun, kondisi ini dapat memburuk seiring waktu dan sering kambuh, sehingga secara signifikan membatasi kemampuan individu dalam menjalani aktivitas sehari-hari. Pada penelitian ini, lima algoritma clustering dibandingkan, yaitu Single Linkage, Average Linkage, Complete Linkage, Centroid Linkage, dan ward Linkage. Dari kelima algoritma tersebut, hasil clustering yang paling optimal adalah Complete Linkage. Hal ini ditunjukkan oleh nilai Chopenetic Correlation yang sebesar 0.726, yang menunjukkan bahwa algoritma ini menghasilkan pengelompokan yang paling akurat dan stabil dalam menggambarkan analisis data. Untuk menentukan jumlah cluster yang optimal, digunakan perhitungan Davies-Bouldin Index (DBI). Nilai DBI yang lebih rendah menunjukkan clustering yang lebih baik. Dalam analisis ini, diperoleh bahwa jumlah cluster yang optimal adalah 4, dengan nilai DBI terendah sebesar 1.171. Ini menunjukkan bahwa empat cluster ini adalah yang paling baik dalam memisahkan dan mengelompokkan data secara efektif. Dari proses AHC ini dari 1030 data didapatkan 4 cluster dengan anggota cluster terbanyak berada pada cluster 2 (cluster dengan kondisi gejala depresi ringan) dengan presentase 32.5%, lalu anggota cluster terbanyak kedua adalah cluster 4 (cluster dengan kondisi gejala depresi berat) dengan persentase 28.3%, kemudian cluster 3 (cluster dengan kondisi gejala depresi sedang) dengan persentase 21.9%, dan cluster 1 (cluster dengan kondisi normal) dengan persentase 17.2%..

#### DAFTAR REFERENSI

- Aan, A., Permana, J., Putu, N., & Puspa, N. (2024). Pengumpulan data tweet berdasarkan kata kunci depresi dan kisah hidup di kalangan mahasiswa berbasis PHQ-9. [*Nama Jurnal Tidak Dicantumkan*], 21(1), 24–33.
- Dirgayunita, A. (2016). Depresi: Ciri, penyebab dan penanganannya. *Jurnal An-Nafs: Kajian dan Penelitian Psikologi*, 1(1), 1–14. <https://doi.org/10.33367/psi.v1i1.235>
- Fatimah, N. (2014). *Desain studi kasus*. Fakultas Kedokteran UIN Jakarta.
- Firdaus, R. D., Laksana, T. G., & Ramadhani, R. D. (2019). Pengelompokan data persediaan obat menggunakan perbandingan metode K-means dengan hierarchical clustering single linkage. *Jurnal Informatics, Information System, Software Engineering and Applications*, 2(1), 33–48.
- Hendra Perdana, N. A. N. S. (2019). Pencarian cluster optimum pada single linkage, complete linkage dan average linkage. *Bimaster: Buletin Ilmiah Matematika, Statistik dan Terapannya*, 8(3), 393–398. <https://doi.org/10.26418/bbimst.v8i3.33173>

- Indriyaningrum, N. B., Brahmantyo, H. P., & Mustika, S. (2017). Derajat depresi pasien hepatitis C kronis yang mendapat terapi PegIFN- $\alpha$ . *Jurnal Kedokteran Brawijaya*, 29(3), 234–237. <https://doi.org/10.21776/ub.jkb.2017.029.03.9>
- Kartikasari, M. D. (2021). Self-organizing map menggunakan Davies-Bouldin Index dalam pengelompokan wilayah Indonesia berdasarkan konsumsi pangan. *Jambura Journal of Mathematics*, 3(2), 187–196. <https://doi.org/10.34312/jjom.v3i2.10942>
- Maulana, M. R., & Al Idrus, S. I. (2023). Sistem pakar untuk mengukur tingkat depresi mahasiswa menggunakan metode Fuzzy Sugeno. *Ocean Engineering: Jurnal Ilmu Teknik dan Teknologi Maritim*, 2(1), 37–50.
- Miharja, M., & Adhkar, S. (2022). Implementasi chatbot deteksi depresi dini pada mahasiswa dengan PHQ-9 (Patient Health Questionnaire) menggunakan NLP (Natural Language Processing). *Prosiding Saintek*, 1(1), 103–108.
- Prayitno, E., Tarigan, N., Sukmawaty, W., & Maudzoh, U. (2022). Gangguan mental emosional dan depresi pada remaja. *Kebangkitan UMKM Pascapandemi Covid-19*, 2(4), 4787–4794. <https://www.bajangjournal.com/index.php/J-ABDI/article/view/3641/2684>
- Seino, Y., et al. (2018). A cluster analysis of bronchial asthma patients with depressive symptoms. *Internal Medicine*, 57(14), 1967–1975. <https://doi.org/10.2169/internalmedicine.9073-17>
- Senjaya, A. A., et al. (2023). Increasing awareness of depression in adolescents and children through online campaign activities. *Jurnal Layanan Masyarakat (Journal of Public Service)*, 7(3), 326–331. <https://doi.org/10.20473/jlm.v7i3.2023.326-331>
- 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>
- Tamara, R. (2023). Data mining penentuan jurusan siswa menggunakan metode agglomerative hierarchical clustering (AHC). *Jurnal Media Informatika Budidarma*, 7(2), 873–880. <https://doi.org/10.30865/mib.v7i2.6092>
- Wijaya, A. E., Asmin, E., & Sapteno, L. B. E. (2023). Tingkat depresi dan ansietas pada usia produktif. *Jurnal Ilmiah Kesehatan Sandi Husada*, 12(1), 150–156.
- Wulandari, S. (2023). Clustering Indonesian provinces on prevalence of stunting toddlers using agglomerative hierarchical clustering. *Faktorexacta*, 16(2). <https://doi.org/10.30998/faktorexacta.v16i2.17186>
- Yoduke, F., Daulima, N. H., & Mustikasari, M. (2023). Peran guru PAI dalam bimbingan dan konseling terhadap pembentukan akhlak siswa di sekolah dasar. *Alauddin Scientific Journal of Nursing*, 4(1), 16–24.
- Yusuf, M., & Wahyu, A. (2022). Analysis of the grouping of provinces in Indonesia according to the democracy index with the agglomerative hierarchical clustering algorithm. *[Nama Jurnal Tidak Dicantumkan]*, 15, 27–35.