

ANALISIS POLA PENYEBARAN PENYAKIT MENGGUNAKAN PENDEKATAN CLUSTERING HIERARKIS DAN K-MEANS

Dilliana Tugas Setiyawan*¹⁾, Berlilana²⁾, Azhari Shouni Barkah³⁾

1. Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Indonesia
2. Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Indonesia
3. Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Indonesia

Article Info

Kata Kunci: Puskesmas, Penyakit, Clustering, Hirarkis, Kmeans,

Keywords: Health Center, Disease, Clustering, Hirarkis, Kmeans,

Article history:

Received 10 January 2025

Revised 16 February 2025

Accepted 20 March 2025

Available online 2 December 2025

DOI :

<https://doi.org/10.29100/jipi.v10i4.7328>

* Corresponding author.

Corresponding Author

E-mail address:

polibara@gmail.com

ABSTRAK

Penyebaran penyakit, baik yang bersifat menular maupun tidak menular, merupakan isu penting yang harus diidentifikasi secara tepat untuk mendukung upaya pencegahan dan pengendalian kesehatan masyarakat. Identifikasi pola sebaran penyakit menjadi krusial karena setiap penyakit memiliki karakteristik penyebaran yang berbeda, baik berdasarkan faktor lingkungan, demografi, maupun perilaku masyarakat. Penerapan K-Means Cluster Analysis merupakan metode yang digunakan untuk mengelompokkan data menjadi beberapa kelompok (cluster) berdasarkan kesamaan karakteristik. Selain itu, pendekatan *Hierarchical Clustering* diterapkan untuk memvisualisasikan hubungan antar data secara hierarkis, memungkinkan analisis yang lebih mendalam. Tujuan penelitian ini adalah untuk menganalisis pola penyebaran penyakit menggunakan pendekatan Clustering Hierarkis dan K-Means. Data dari 35 Puskesmas dianalisis berdasarkan jumlah pasien dan prevalensi penyakit, termasuk Tuberkulosis, Diabetes, Hipertensi, dan penyakit menular lainnya. Hasil penelitian menunjukkan bahwa kedua metode memberikan wawasan yang saling melengkapi. K-Means efektif dalam membagi data menjadi cluster yang merata dan efisien, sementara *Hierarchical Clustering* memungkinkan identifikasi hubungan hierarkis dan distribusi granular antar Puskesmas

ABSTRACT

The spread of diseases, both communicable and non-communicable, is an important issue that must be properly identified to support public health prevention and control efforts. Identification of disease distribution patterns is crucial because each disease has different distribution characteristics, both based on environmental factors, demographics, and community behavior. The application of K-Means Cluster Analysis is a method used to group data into several groups (clusters) based on similar characteristics. In addition, the Hierarchical Clustering approach is applied to visualize the relationship between data hierarchically, allowing for more in-depth analysis. The purpose of this study was to analyze disease spread patterns using Hierarchical Clustering and K-Means approaches. Data from 35 Puskesmas were analyzed based on the number of patients and prevalence of diseases, including Tuberculosis, Diabetes, Hypertension, and other infectious diseases. The results showed that both methods provided complementary insights. K-Means was effective in dividing the data into evenly distributed and efficient clusters, while Hierarchical Clustering enabled the identification of hierarchical relationships and granular distributions between health centers.

I. PENDAHULUAN

Penyebaran penyakit, baik yang bersifat menular maupun tidak menular, merupakan isu penting yang harus diidentifikasi secara tepat untuk mendukung upaya pencegahan dan pengendalian kesehatan masyarakat. Identifikasi pola sebaran penyakit menjadi krusial karena setiap penyakit memiliki karakteristik penyebaran yang berbeda, baik berdasarkan faktor lingkungan, demografi, maupun perilaku masyarakat [1]. Provinsi Jawa Tengah mengalami peningkatan pada penyakit tidak menular seperti kejadian stroke dan penyakit

jantung iskemik sebagai penyebab kematian tertinggi dan penyakit menular seperti tuberkulosis dan infeksi saluran pernapasan juga mengalami peningkatan terutama di wilayah dengan akses kesehatan yang terbatas [2].

Kabupaten Banjarnegara, yang terletak di Provinsi Jawa Tengah, merupakan daerah dengan kondisi geografis yang unik dan tantangan sanitasi yang signifikan. Sebagai salah satu kabupaten dengan lebih dari 60% penduduknya tinggal di daerah pegunungan, Banjarnegara memiliki topografi yang sangat beragam, mulai dari kawasan dataran rendah hingga perbukitan yang curam. Keadaan geografis ini memberikan tantangan tersendiri dalam hal aksesibilitas dan penyediaan layanan kesehatan, terutama di daerah-daerah pegunungan yang terpencil. Puskesmas yang tersebar di seluruh wilayah Kabupaten, yang berjumlah 35, menjadi ujung tombak dalam memberikan layanan kesehatan kepada masyarakat. Faktor topografi yang sulit dan keterbatasan infrastruktur seringkali menghambat distribusi layanan kesehatan yang merata. Selain itu, faktor iklim yang cukup ekstrem juga berkontribusi terhadap dinamika kesehatan masyarakat di Banjarnegara. Curah hujan yang tinggi, terutama selama musim penghujan, dapat menyebabkan banjir dan penyakit yang terkait dengan air, seperti diare dan demam berdarah, sementara suhu yang lebih rendah di daerah pegunungan dapat meningkatkan risiko penyakit pernapasan. Dari segi ekonomi, meskipun sebagian besar penduduknya bergantung pada sektor pertanian, ketergantungan pada kegiatan ekonomi tradisional ini juga berhubungan dengan masalah sanitasi yang buruk di beberapa wilayah, memperburuk penyebaran penyakit. Kondisi ini semakin diperparah dengan terbatasnya kesadaran masyarakat tentang pentingnya menjaga kebersihan lingkungan dan fasilitas sanitasi yang memadai. Penting untuk memahami faktor-faktor geografis, iklim, dan ekonomi Kabupaten Banjarnegara dalam mengkaji pola penyebaran penyakit, karena ketiga faktor tersebut berinteraksi dan berkontribusi pada tingkat kesehatan masyarakat yang ada. Data yang diperoleh dari Dinas Kesehatan menunjukkan adanya tren peningkatan jumlah penderita berbagai penyakit dalam beberapa tahun terakhir [3]. Sehingga diperlukan identifikasi pola dan subkelompok data penyakit, salah satunya berupa *clustering* sebagai teknik analisis data memungkinkan peneliti untuk mengelompokkan pasien atau kasus penyakit berdasarkan kesamaan karakteristik, yang dapat membantu mendiagnosis, pengobatan, dan pemahaman lebih lanjut tentang penyakit tertentu [4].

Pola penyebaran penyakit terhadap kelompok rentan yang lebih berisiko, seperti anak-anak, lansia, atau individu dengan kondisi kesehatan tertentu. Selain itu, faktor wilayah juga memainkan peran penting dalam penyebaran penyakit, karena perbedaan geografis, kondisi sanitasi, dan akses terhadap layanan kesehatan dapat mempengaruhi tingkat penularan dan prevalensi penyakit. Diperlukan pendekatan berbasis wilayah untuk mengidentifikasi daerah - daerah dengan tingkat risiko tinggi, sehingga langkah-langkah preventif dan pengobatan dapat difokuskan pada lokasi dan kelompok yang tepat, baik untuk penyakit menular seperti demam berdarah maupun penyakit tidak menular seperti diabetes [1].

Penyakit jantung iskemik menempati urutan pertama sebagai penyebab kematian terbesar sekitar 13% dari total kematian di seluruh dunia pada Tahun 2021. COVID-19 muncul sebagai penyebab kematian kedua terbesar sejumlah 8,7 juta kematian pada tahun yang sama. Penyakit lain seperti stroke, penyakit paru obstruktif kronik (COPD), dan infeksi saluran pernapasan bawah menjadi penyebab kematian terbesar di dunia [5].

Kejadian penyakit di negara Asia meliputi penyakit stroke, jantung iskemik, diabetes, dan penyakit paru obstruktif kronik menjadi penyebab utama kematian. Kejadian COVID-19 pada tahun 2020 telah memperparah masalah kesehatan di banyak negara Asia, dengan infeksi saluran pernapasan seperti pneumonia dan penyakit paru-paru lainnya menjadi masalah kesehatan yang signifikan [6]. Sebaran penyakit di Indonesia, stroke merupakan penyebab kematian tertinggi, menyumbang 21,2% dari total kematian sedangkan penyakit jantung iskemik berada di peringkat kedua dan diikuti diabetes, infeksi saluran pernapasan dan tuberkulosis terdaftar dalam 10 besar penyakit penyebab kematian [7].

Penelitian sebelumnya, seperti yang dilakukan oleh Bisquera (2021), lebih berfokus pada identifikasi pola penyakit di lingkungan perkotaan, tanpa memperhatikan karakteristik wilayah rural yang memiliki akses kesehatan terbatas, seperti Kabupaten Banjarnegara. Hal ini menciptakan kesenjangan dalam memahami pola penyakit di daerah dengan tantangan geografis seperti daerah pegunungan. Penelitian Azmiyannoor (2023) mengidentifikasi faktor-faktor yang menghambat pelaksanaan pemetaan penyakit. Tidak ada penekanan pada integrasi metode clustering untuk analisis gabungan penyakit menular dan tidak menular. Hal ini penting mengingat kedua jenis penyakit sering saling memengaruhi dalam konteks masyarakat.

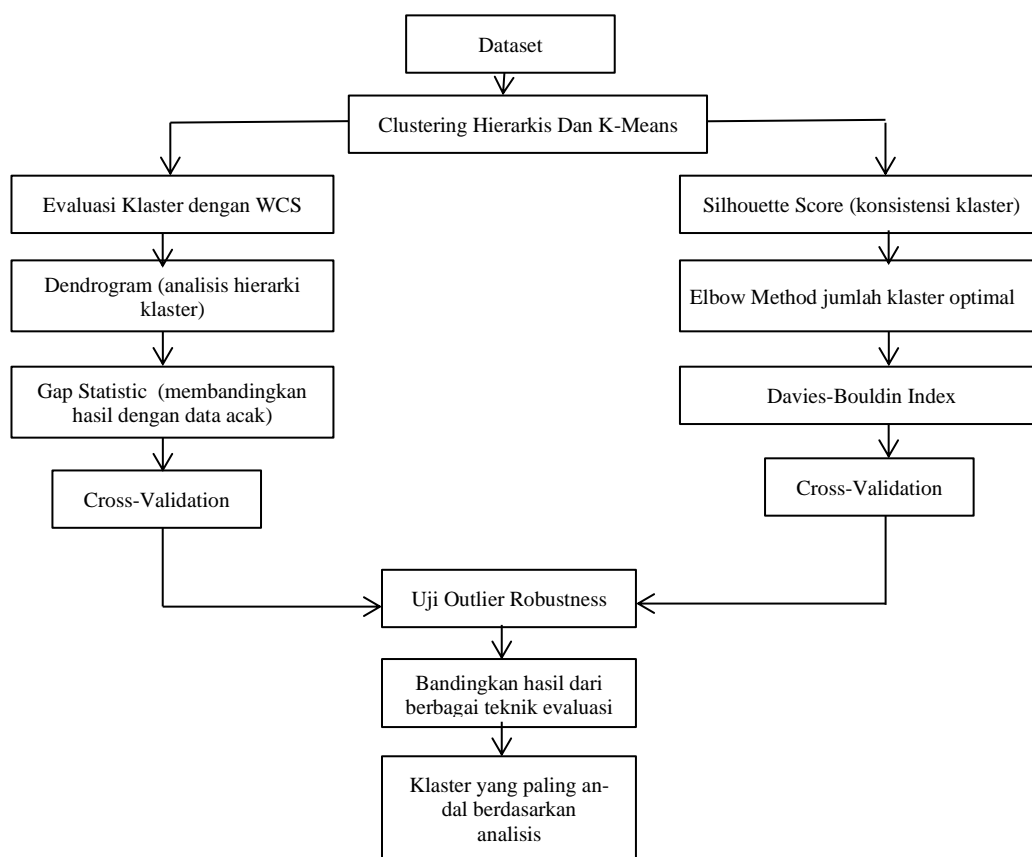
Penggunaan metode K-Means penting dalam konteks analisis kesehatan dan epidemiologi. Kebutuhan untuk mengidentifikasi pola dan subkelompok dalam data penyakit menjadi semakin mendesak. *Clustering* sebagai teknik analisis data memungkinkan peneliti untuk mengelompokkan pasien atau kasus penyakit berdasarkan kesamaan karakteristik, yang dapat membantu dalam diagnosis, pengobatan, dan pemahaman lebih lanjut tentang penyakit tertentu [4].

Jumlah penduduk Kabupaten Banjarnegara pada tahun 2024 tercatat sebanyak 1.061.258 jiwa, dengan sarana kesehatan tingkat pertama yang terdiri dari 35 Puskesmas. Total kunjungan masyarakat yang berobat di Puskesmas sepanjang tahun 2024 mencapai 854.405 kali. Tingginya jumlah kunjungan di Puskesmas, menjadi sangat penting

untuk mengidentifikasi pola penyebaran penyakit di wilayah tersebut secara lebih terstruktur dan efisien. Belum ada studi yang secara komprehensif membandingkan performa metode clustering hierarkis dan K-Means dalam konteks epidemiologi, khususnya di wilayah dengan karakteristik unik seperti Kabupaten Banjarnegara. Sebagian besar penelitian fokus pada lingkungan perkotaan, sementara konteks daerah pegunungan dengan akses kesehatan terbatas, seperti di Banjarnegara, belum banyak dieksplorasi. memanfaatkan data berbasis lokasi untuk mendukung identifikasi kluster penyakit secara dinamis, yang dapat membantu pengambilan keputusan berbasis data untuk distribusi layanan kesehatan. Tantangan ini menciptakan kebutuhan untuk memahami pola distribusi penyakit yang spesifik untuk mendukung strategi pencegahan yang lebih efektif. Pada penelitian ini akan dilakukan analisis pola penyebaran penyakit menggunakan pendekatan clustering hierarkis dan k-means di Kabupaten Banjarnegara pada tahun 2024.

II. METODE PENELITIAN

Penelitian ini menggunakan data Laporan 35 Puskesmas di Kabupaten Banjarnegara Propinsi Jawa Tengah dari bulan Januari – Desember 2023. Metode penelitian terlihat pada gambar berikut



Gambar 1. Diagram Clustering Hierarkis Dan K-Means

A. Pengumpulan data proses

Proses pengumpulan data kesehatan di Kabupaten Banjarnegara dilakukan oleh Dinas Kesehatan Kabupaten melalui sistem pelaporan yang diterima dari seluruh puskesmas di wilayah tersebut. Setiap puskesmas secara rutin mengirimkan laporan yang mencakup berbagai informasi penting mengenai kondisi kesehatan masyarakat Data yang dikumpulkan meliputi jumlah individu yang melakukan pemeriksaan kesehatan, yang dibedakan berdasarkan jenis kelamin. Selain itu, puskesmas juga melaporkan jumlah masyarakat yang terdiagnosis dengan berbagai penyakit menular dan tidak menular, seperti Tuberkulosis, HIV, Hepatitis, Kusta, Demam Berdarah, Malaria, Hipertensi, Diabetes, dan Kanker Rahim. Laporan-laporan ini kemudian dianalisis untuk memantau tren kesehatan masyarakat, merencanakan program intervensi yang tepat, dan mengambil langkah-langkah preventif guna meningkatkan kualitas kesehatan di wilayah Kabupaten Banjarnegara Jawa Tengah. Tabel distribusi sebaran penyakit dapat dilihat pada tabel 1

TABEL I
 HASIL DISTRIBUSI PENYAKIT

No	Puskesmas	Pasien puskesmas	Tuberkulosis	Demam berdarah	Diabetes	Hipertensi	HIV	Kusta	Malaria	Kanker Rahim
1	Susukan 1	20.343	238	2	620	9500	4	4	25	2.575
2	Susukan 2	9.525	144	0	351	4856	2	0	0	2.100
3	Klampok 1 Pwj	14.723	121	2	508	6881	1	1	16	1.985
4	Klampok 2	7.379	84	1	253	4354	5	0	0	2.564
5	Mandiraja 1	14.832	343	12	717	10687	6	1	202	6.143
6	Mandiraja 2	17.496	286	15	621	9662	4	0	15	7.079
7	Purwonegoro 1	17.530	268	13	573	10965	8	1	9	5.509
8	Purwonegoro 2	18.709	205	15	703	10524	2	0	6	6.280
9	Bawang 1	16.823	462	7	618	10465	9	0	2	6.297
10	Bawang 2	12.570	301	2	411	6612	5	0	3	1.621
11	Banjarnegara 1	20.540	298	8	715	9715	14	0	0	2.229
12	Banjarnegara 2	17.661	289	8	337	7525	8	0	1	9.430
13	Pagedongan	14.481	311	7	637	9551	2	0	518	6.043
14	Sigaluh 1	9.582	175	4	323	4725	2	0	20	2.773
15	Sigaluh 2	6.654	159	0	218	3565	2	0	8	2.248
16	Madukara 1	9.632	186	1	323	5365	0	0	0	3.109
17	Madukara 2	13.035	91	3	392	6962	2	0	0	3.991
18	Banjarmangu 1	10.125	254	7	377	6123	7	0	751	3.529
19	Banjarmangu 2	8.900	148	2	149	4852	4	0	510	3.247
20	Wanadadi 1	10.115	166	2	311	5165	2	1	28	3.041
21	Wanadadi 2	8.269	150	6	251	3465	2	0	0	2.247
22	Rakit 1	15.766	355	14	503	8852	12	0	0	4.459
23	Rakit 2	14.197	170	6	337	5462	4	0	94	3.340
24	Punggelan 1	21.719	276	5	718	10335	11	1	100	7.423
25	Punggelan 2	19.148	104	2	663	8687	6	0	66	6.341
26	Karangkobar	15.266	540	2	575	8254	0	1	8	7.184
27	Pagentan 1	10.483	138	2	376	5941	6	0	0	3.635
28	Pagentan 2	7.405	75	6	224	3862	2	0	0	3.743
29	Pejawaran	18.374	466	3	531	7752	5	0	0	5.693
30	Batur 1	10.569	78	0	335	5336	2	0	0	3.647
31	Batur 2	6.792	47	0	221	3416	0	0	0	2.215
32	Wanayasa 1	11.459	148	3	438	6642	4	0	2	4.024
33	Wanayasa 2	10.993	226	0	419	5742	4	0	0	3.198
34	Kalibening	16.192	507	2	649	9952	4	2	0	7.534
35	Pandanarum	10.153	289	0	368	5723	5	1	0	1.444

Sumber : Data Laporan Tahunan Dinas Kesehatan Kabupaten Banjarnegara Tahun 2024

B. Cluster data

Cluster data berupa pengelompokan data berdasarkan kemiripan atau kesamaan karakteristik tertentu [8]. Pada konteks *machine learning*, clustering adalah metode *unsupervised learning* yang digunakan untuk menemukan pola atau struktur dalam data tanpa memerlukan label atau kategori yang telah ditentukan sebelumnya [9].

1) Pengklusteran K-Means

K-Means adalah salah satu algoritma pengelompokan (clustering) yang paling umum digunakan dalam analisis data. Algoritma ini bekerja dengan membagi data menjadi beberapa kelompok (klaster) berdasarkan kesamaan karakteristik antar data [10]. Proses K-Means dimulai dengan menentukan jumlah klaster (k) yang diinginkan dan menginisialisasi pusat klaster (*centroid*) secara acak. Kemudian, algoritma mengalokasikan setiap data ke klaster terdekat berdasarkan jarak Euclidean. Setelah itu, pusat klaster diperbarui dengan menghitung rata-rata posisi data dalam klaster. Proses ini berulang hingga pusat klaster stabil, yaitu ketika tidak ada perubahan signifikan dalam posisi pusat klaster atau ketika iterasi mencapai batas maksimum [11]. Kelebihan K-Means berupa kemudahan implementasi dan kecepatan komputasi, namun algoritma ini sensitif terhadap pemilihan jumlah klaster dan posisi awal pusat klaster [12].

2) Hierarchical Clustering

Hierarchical Clustering berupa metode pengelompokan data yang membangun struktur hierarkis untuk menunjukkan hubungan antar data tanpa memerlukan penentuan jumlah cluster di awal. Metode ini terdiri dari dua pendekatan utama, yaitu Agglomerative (*Bottom-Up*) dan Divisive (*Top-Down*) [13]. Pada pendekatan *agglomerative*, setiap data dimulai sebagai cluster individu yang digabungkan secara bertahap

hingga menjadi satu cluster besar, sedangkan pada pendekatan divisive, data dimulai dalam satu cluster besar yang kemudian dipecah menjadi cluster lebih kecil hingga setiap data menjadi cluster individu [14].

Pengukuran jarak antar cluster dapat dilakukan dengan beberapa metode, seperti Single Linkage (jarak terpendek antara dua titik dalam cluster yang berbeda), Complete Linkage (jarak terjauh antara dua titik), Average Linkage (rata-rata jarak antar semua pasangan titik), dan Ward's Method (meminimalkan jumlah kuadrat total dalam cluster). Hasil dari Hierarchical Clustering biasanya divisualisasikan dalam bentuk dendrogram, yang menunjukkan bagaimana data digabungkan atau dipecah pada berbagai tingkat hierarki. Pemotongan dendrogram pada tingkat tertentu memungkinkan penentuan jumlah cluster yang dihasilkan[15]

3) Evaluasi Hasil Clustering K-Means dan Hierarchical Clustering

Tahap evaluasi ini bertujuan untuk mengukur kinerja dari model K-Means dan Hierarchical Clustering yang telah diterapkan pada dataset yang berisi data pasien Tuberkulosis, Diabetes, Hipertensi, demam berdarah, HIV, Kusta, Malaria dan Kanker Rahim, Evaluasi model K-Means dan Hierarchical Clustering dilakukan dengan membagi data pasien menjadi 80% untuk pelatihan dan 20% untuk pengujian guna menghindari overfitting. *Confusion Matrix* digunakan untuk menilai kinerja model melalui metrik seperti Akurasi, Presisi, Recall, dan F1-Score.

a) Akurasi

Akurasi mengukur seberapa baik model klasifikasi membuat prediksi yang benar, dihitung sebagai rasio prediksi benar (True Positive + True Negative) terhadap total prediksi

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}$$

b) Presisi

Presisi mengukur akurasi dari prediksi positif, yaitu proporsi prediksi positif yang benar-benar positif

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

c) Recall

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

d) F1 Score

$$\text{F1Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

TABEL II
 HASIL *CONFUSION MATRIX*

	Hierarchical Clustering	K-Means
Metric	Score	Score
Accuracy	0.514286	0.486
Precision	0.500000	0.467
Recall	0.529412	0.412
F1-Score	0.514286	0.438

Berdasarkan pada tabel II menunjukkan K-Means memiliki akurasi 48.6%, presisi 47.8%, recall 64.7%, dan F1-Score 55%, mencerminkan kinerja moderat dengan ruang untuk perbaikan. Hierarchical Clustering lebih unggul dengan akurasi 51.4%, presisi 50%, recall 52.9%, dan F1-Score 51.4%, menunjukkan keseimbangan yang lebih baik dalam deteksi pola data. K-Means lebih efisien untuk pengelompokan, sementara Hierarchical Clustering unggul dalam mendeteksi hubungan antar data. Keduanya berpotensi membantu analisis distribusi penyakit di Puskesmas untuk mendukung pengambilan keputusan kesehatan berbasis data.

III. HASIL DAN PEMBAHASAN

Data dari 35 Puskesmas mencakup berbagai variabel penting yang digunakan untuk analisis yang meliputi nama Puskesmas sebagai identitas dari setiap fasilitas kesehatan, jumlah total pasien, yang kemudian dirinci lebih lanjut menjadi jumlah pasien laki-laki dan jumlah pasien perempuan. Data juga mencakup jumlah pasien yang terdiagnosis penyakit tertentu, seperti Tuberkulosis, yang dibagi menjadi total kasus, jumlah pasien laki-laki, dan perempuan. Hal serupa juga berlaku untuk penyakit lainnya, seperti

Demam Berdarah, Diabetes, Hipertensi, dan HIV, yang mencatat total kasus serta pembagian berdasarkan jenis kelamin.

Selain itu, terdapat data mengenai Jumlah pasien Kusta dan Jumlah pasien Malaria, yang masing-masing juga dirinci berdasarkan jenis kelamin. Variabel lainnya, seperti skrining deteksi kanker perempuan, menunjukkan jumlah pasien yang mengikuti pemeriksaan kanker, serta jumlah pasien yang melakukan pemeriksaan lanjutan atau tidak melanjutkan pemeriksaan. Seluruh variabel ini memberikan gambaran komprehensif mengenai kondisi kesehatan masyarakat di setiap Puskesmas dan menjadi dasar untuk melakukan analisis lebih lanjut.

TABEL III
 HASIL DISTRIBUSI CLUSTER

No	PUSKESMAS	Cluster 2	Cluster 3	Cluster 4	Cluster 5
1	Susukan 1	0	1	1	1
2	Susukan 2	1	2	0	0
3	Pwj Klampok 1	1	0	3	3
4	Pwj Klampok 2	1	2	2	2
5	Mandiraja 1	0	1	1	4
6	Mandiraja 2	0	1	1	1
7	Purwonegoro 1	0	1	1	4
8	Purwonegoro 2	0	1	1	4
9	Bawang 1	0	1	1	4
10	Bawang 2	1	0	3	3
11	Banjarnegara 1	0	1	1	1
12	Banjarnegara 2	0	0	3	3
13	Pagedongan	0	1	1	1
14	Sigaluh 1	1	2	0	0
15	Sigaluh 2	1	2	2	2
16	Madukara 1	1	2	0	0
17	Madukara 2	1	0	3	3
18	Banjarmangu 1	1	0	0	3
19	Banjarmangu 2	1	2	0	0
20	Wanadadi 1	1	2	0	0
21	Wanadadi 2	1	2	2	2
22	Rakit 1	0	1	1	1
23	Rakit 2	1	2	0	0
24	Punggelan 1	0	1	1	4
25	Punggelan 2	0	1	1	1
26	Karangkoobar	0	0	3	1
27	Pagentan 1	1	0	0	0
28	Pagentan 2	1	2	2	2
29	Pejawaran	0	0	3	3
30	Batur 1	1	2	0	0
31	Batur 2	1	2	2	2
32	Wanayasa 1	1	0	3	3
33	Wanayasa 2	1	2	0	0
34	Kalibening	0	1	1	4
35	Pandanarum	1	2	0	0

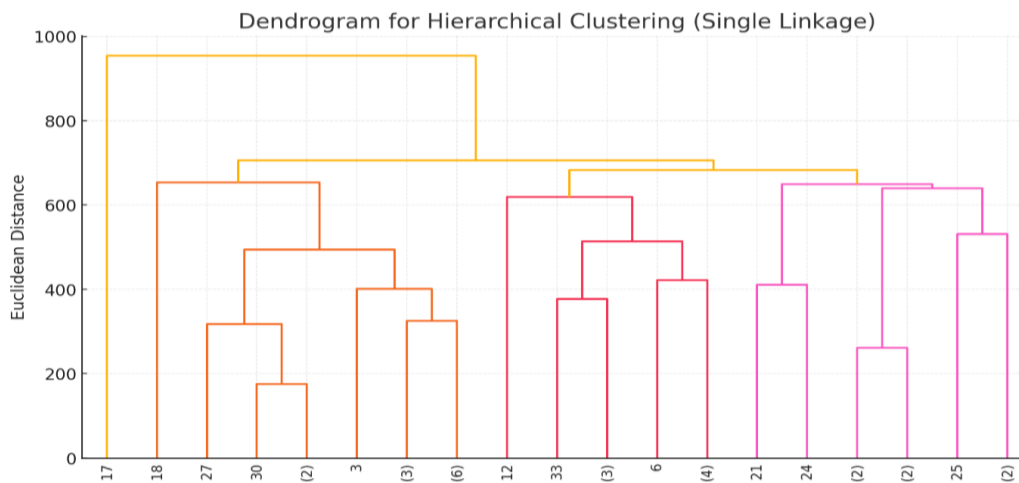
Berdasarkan tabel tiga di atas didapatkan hasil bagaimana data dari 35 Puskesmas dikelompokkan berdasarkan pola distribusi jumlah pasien dan jenis penyakit yang dilaporkan. Pada Cluster 2, data dikelompokkan menjadi dua kategori besar yang mencerminkan perbedaan signifikan, seperti Puskesmas dengan jumlah pasien tinggi dan rendah. Ketika jumlah cluster ditambah menjadi 3 cluster, pembagian menjadi lebih rinci, dengan identifikasi kelompok menengah yang mengisi celah antara dua kelompok utama. Pada Cluster 4, analisis menghasilkan distribusi yang lebih detail, membagi data berdasarkan karakteristik spesifik, seperti variasi jumlah kasus penyakit tertentu, misalnya Tuberkulosis atau Diabetes. Pada Cluster 5, pengelompokan menjadi semakin granular, di mana setiap cluster mencerminkan kelompok dengan pola distribusi data yang unik. Peningkatan jumlah cluster memberikan kejelasan yang lebih besar tentang perbedaan antara Puskesmas, tetapi juga meningkatkan kompleksitas analisis. Cluster dengan jumlah lebih kecil, seperti 2 atau 3, lebih cocok untuk analisis umum, sedangkan cluster yang lebih besar, seperti 4 atau 5, berguna untuk memahami pola spesifik yang dapat mendukung pengambilan keputusan kesehatan yang lebih mendetail [16]. Kombinasi analisis ini memberikan pandangan menyeluruh tentang distribusi dan karakteristik Puskesmas berdasarkan data yang tersedia sebagai perencanaan dan pengelolaan layanan kesehatan yang lebih efektif. Hasil *Davies-Bouldin Index Results* sebagai berikut :

TABEL IV
 HASIL DBI

Cluster	DBI
2	0.439468
3	0.572643
4	0.511800
5	0.591107

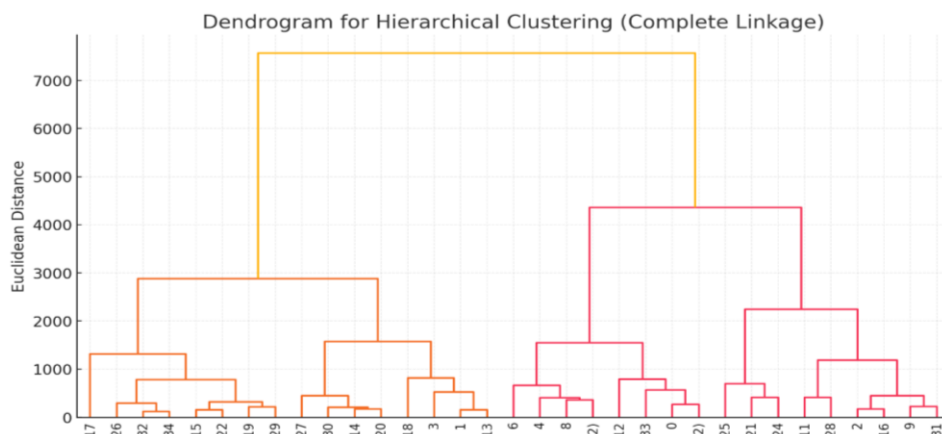
Berdasarkan tabel IV didapati hasil analisis *Davies-Bouldin Index* (DBI) menunjukkan bahwa kualitas clustering bervariasi tergantung pada jumlah cluster yang digunakan. Pada Cluster 2, nilai DBI sebesar 0.439 adalah yang paling rendah, mengindikasikan bahwa pembagian data menjadi dua cluster memiliki pemisahan terbaik. Hal ini sejalan dengan temuan yang diungkapkan oleh Nowak bahwa pemilihan jumlah cluster yang tepat sangat penting untuk mencapai hasil clustering yang optimal [17].

Ketika jumlah cluster ditambah menjadi Cluster 3, nilai DBI meningkat menjadi 0.573, menunjukkan bahwa kualitas clustering menurun. Cluster ini mulai kehilangan homogenitas dalam kelompoknya. Pada Cluster 4, nilai DBI menurun menjadi 0.512, menandakan sedikit perbaikan dibandingkan cluster 3, untuk Cluster 5 memiliki nilai DBI tertinggi, yaitu 0.591, menunjukkan bahwa clustering ini kurang mampu mempertahankan pemisahan yang jelas antar cluster. Hasil cluster yang lebih banyak tidak selalu menjamin pemisahan yang lebih baik [18]. Secara keseluruhan, hasil ini mengindikasikan bahwa pembagian data menjadi 2 cluster merupakan pilihan terbaik untuk mendapatkan hasil clustering yang lebih optimal dan seragam [19]. Hasil Implementasi pengelompokan hirarki Aglomeratif sebagai berikut :



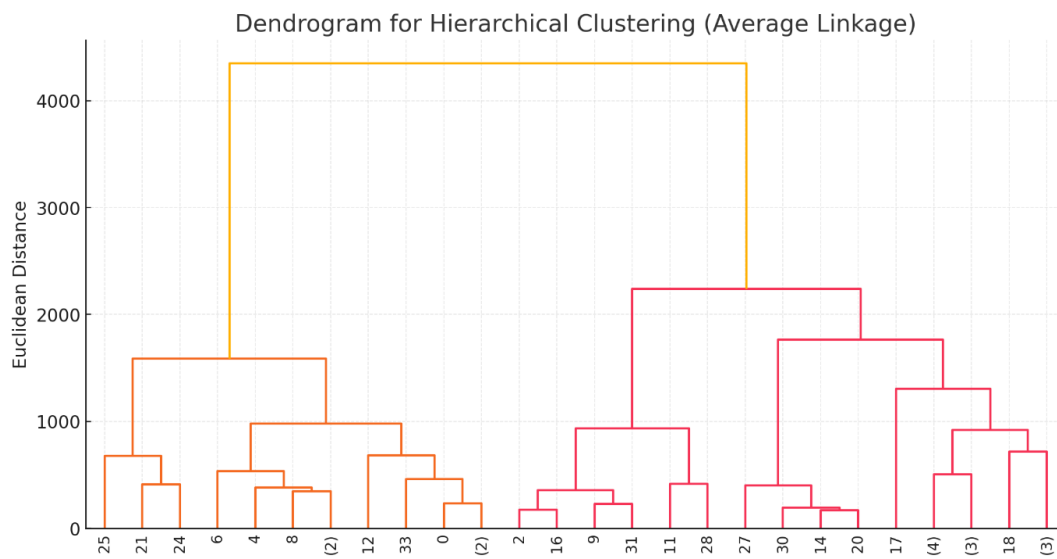
Gambar 1. Single linkage

Berdasarkan gambar 1 didapati metode *Single Linkage*, pengelompokan didasarkan pada jarak terpendek antar data. Hal ini menghasilkan cluster yang terhubung secara bertahap, tetapi sering kali memunculkan efek chaining, di mana cluster menjadi terlalu panjang dan tidak kompak. Dalam kasus ini, Puskesmas seperti Sigaluh 2 dan Pwj Klampok 2, yang memiliki jumlah pasien relatif rendah, cenderung terhubung dalam rantai yang sama dengan Puskesmas dengan jumlah pasien lebih tinggi. Metode *Single Linkage* dapat menghasilkan hasil clustering yang baik, tetapi juga mengakui potensi masalah chaining yang dapat muncul dalam pengelompokan data [20].



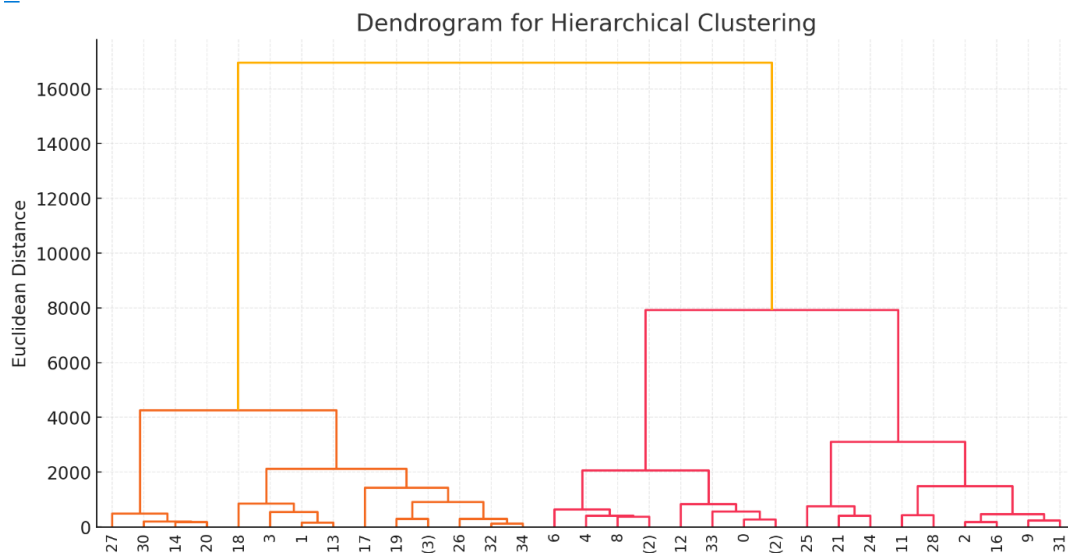
Gambar 2. Dendrogram dari metode tautan lengkap

Berdasarkan gambar 2 bahwa metode *Complete Linkage* mengelompokkan data berdasarkan jarak maksimum antar data, menghasilkan cluster yang lebih kompak dan terpisah. Dalam kasus ini, Puskesmas dengan jumlah pasien besar, seperti Susukan 1 dan Punggelan 1, lebih jelas terpisah dari Puskesmas lainnya. Hal ini memudahkan dalam mengidentifikasi kelompok dengan prevalensi penyakit yang tinggi, seperti Tuberkulosis dan Diabetes. Secara keseluruhan, penggunaan metode *Complete Linkage* dalam analisis clustering Puskesmas memberikan keuntungan signifikan dalam mengidentifikasi kelompok dengan prevalensi penyakit yang tinggi, serta memastikan bahwa cluster yang dihasilkan lebih kompak dan terpisah dengan jelas. Efektivitas metode *Complete Linkage* dalam mengelompokkan data dengan karakteristik serupa dapat dihubungkan dengan konteks pengelompokan Puskesmas berdasarkan jumlah pasien [21].



Gambar 3. dendrogram dari metode hubungan rata-rata

Berdasarkan gambar 3 didapati *Average Linkage* menggunakan rata-rata jarak antar data, memberikan hasil yang seimbang antara *Single* dan *Complete Linkage*. Pada kasus ini, distribusi cluster menjadi lebih merata, memungkinkan analisis yang lebih holistik terhadap hubungan antar Puskesmas. Puskesmas dengan jumlah pasien sedang, seperti Pwj Klampok 1, dapat terkelompokkan dengan baik, mencerminkan distribusi penyakit yang moderat. Penggunaan teknik clustering, termasuk *Average Linkage*, sangat penting dalam sistem kesehatan untuk mengidentifikasi pola dan karakteristik pasien. Dengan mengelompokkan pasien berdasarkan kesamaan dalam data kesehatan, sistem kesehatan dapat lebih efektif dalam merancang intervensi yang sesuai [22].



Gambar 4. Kluster Dendrogram

Berdasarkan gambar 4 didapati dendrogram pada tingkat jarak tertentu, jumlah cluster yang optimal dapat diidentifikasi untuk analisis lebih lanjut. Secara keseluruhan, hasil Hierarchical Clustering memberikan pandangan yang mendalam tentang struktur hubungan antar Puskesmas. Identifikasi kelompok yang relevan untuk mendukung pengambilan keputusan kesehatan secara lebih efektif. Pemotongan dendrogram dapat digunakan untuk menentukan jumlah cluster yang optimal dalam konteks data kesehatan [23].

Hasil analisis clustering menggunakan metode K-Means dan Hierarchical Clustering pada data dari 35 Puskesmas menunjukkan pola distribusi penyakit yang bervariasi terhadap jumlah cluster yang digunakan, dengan jumlah cluster yang sama, kedua metode menghasilkan pembagian data yang berbeda, mencerminkan pendekatan unik masing-masing metode dalam mengelompokkan data. Pada K-Means Clustering, data dikelompokkan berdasarkan minimisasi jarak Euclidean, menghasilkan cluster yang lebih kompak dan terfokus. K-Means efektif dalam mengelompokkan data dengan jumlah besar, dan dapat dengan jelas memisahkan Puskesmas dengan jumlah pasien tinggi dalam satu cluster, sementara Puskesmas dengan jumlah pasien lebih sedikit berada di cluster lainnya [24]. Penelitian ini juga menekankan bahwa dengan peningkatan jumlah cluster, pola distribusi menjadi lebih granular, memungkinkan identifikasi karakteristik spesifik seperti prevalensi penyakit tertentu [24].

Pada cluster dua, Puskesmas dengan jumlah pasien yang tinggi cenderung masuk dalam satu cluster yang sama, sementara Puskesmas dengan jumlah pasien lebih sedikit berada di cluster lainnya. Namun, dengan peningkatan jumlah cluster, pola distribusi menjadi lebih granular, memungkinkan identifikasi karakteristik spesifik seperti prevalensi penyakit tertentu. Hierarchical Clustering menggunakan struktur hierarkis yang memungkinkan analisis hubungan antar Puskesmas secara bertahap. Pada jumlah kluster yang lebih kecil, metode ini menunjukkan hasil yang mirip dengan K-Means dalam memisahkan Puskesmas berdasarkan jumlah pasien. Pada jumlah kluster yang lebih banyak, metode ini memperlihatkan fleksibilitas dalam mengidentifikasi subkelompok berdasarkan kedekatan hierarkis, yang tergambar jelas dalam dendrogram. Hierarchical Clustering dapat memberikan wawasan yang lebih dalam mengenai hubungan antar data, terutama dalam mengidentifikasi subgrup berbasis kedekatan atribut.

Perbandingan antara kedua metode menunjukkan bahwa K-Means lebih efisien untuk analisis dengan jumlah data besar, sementara Hierarchical Clustering lebih baik untuk eksplorasi pola hubungan antar data. Hasil ini memberikan wawasan tentang bagaimana distribusi penyakit seperti Tuberkulosis, Diabetes, dan Hipertensi tersebar di antara Puskesmas dan bagaimana data tersebut dapat digunakan untuk mendukung pengambilan keputusan kesehatan. Penelitian oleh Shah dan Rosli (2023) mendukung temuan ini dengan menunjukkan bahwa analisis data kesehatan menggunakan berbagai metode clustering dapat memberikan wawasan yang lebih baik dalam pengambilan keputusan kesehatan [25].

IV. KESIMPULAN

Hasil analisis clustering menggunakan K-Means dan Hierarchical Clustering menunjukkan pola distribusi penyakit yang signifikan di 35 Puskesmas. Metode K-Means efektif dalam membagi data secara cepat menjadi cluster yang kompak, sementara Hierarchical Clustering memberikan wawasan mendalam melalui struktur hubungan hierarkis antar Puskesmas. K-Means lebih efisien untuk analisis skala besar, tetapi rentan terhadap pemilihan jumlah cluster, sedangkan Hierarchical Clustering unggul dalam fleksibilitas dan visualisasi. Kedua metode ini saling melengkapi, memberikan gambaran makro dan mikro yang bermanfaat untuk pengambilan keputusan kesehatan. Informasi yang diperoleh dapat digunakan untuk merancang strategi yang lebih tepat dalam menangani penyebaran penyakit, mengalokasikan sumber daya, dan meningkatkan layanan kesehatan di Puskesmas yang dianalisis.

V. UCAPAN TERIMAKASIH

Terimakasih kami sampaikan kepada Direktur Politeknik Banjarnegara atas segala dukungan sehingga terselesaikannya penelitian ini.

VI. DAFTAR PUSTAKA

- [1] Kemenkes.RI, "Permenkes RI Nomor 21 Tahun 2020," *Kemeteri. Kesehat. RI*, vol. 9, no. May, p. 6, 2020, [Online]. Available: https://www.slideshare.net/maryamkazemi3/stability-of-colloids%0Ahttps://barnard.edu/sites/default/files/inline/student_user_guide_for_spss.pdf%0Ahttp://www.ibm.com/support%0Ahttp://www.spss.com/sites/dm-book/legacy/ProgDataMgmt_SPSS17.pdf%0Ahttps://www.n
- [2] Dinkes Jateng, "Tengah Tahun 2023 Jawa Tengah," 2023.
- [3] Dinkes Banjarnegara, "Dinas Kesehatan Kabupaten Banjarnegara Tahun 2023," no. 8, 2024.

- [4] I. Landi *et al.*, “Deep representation learning of electronic health records to unlock patient stratification at scale,” *npj Digit. Med.*, vol. 3, no. 1, pp. 1–11, 2020, doi: 10.1038/s41746-020-0301-z.
- [5] WHO, “World Health Organization Global Nutrition Report 2023: Shaping the Future of Global Health. Geneva: WHO. <https://www.who.int/publications/i/item/global-nutrition-report-2023>,” 2023.
- [6] WHO, *World health WORLD HEALTH ORGANIZATION - World health statistics 2024. ISBN 9789240094703. tatistics 2024.* 2024.
- [7] Kementerian Kesehatan RI, “Profil kesehatan Indonesia Tahun 2023,” 2023, [Online]. Available: <https://www.kemkes.go.id/profil-kesehatan-indonesia-2023>
- [8] C. Eckhardt *et al.*, “Unsupervised Machine Learning Methods and Emerging Applications in Healthcare,” *Knee Surg. Sport. Traumatol. Arthrosc.*, vol. 31, no. 2, pp. 376–381, 2022, doi: 10.1007/s00167-022-07233-7.
- [9] Y. Song, X. L. Jia, and H. D. Meng, “Combination Clustering Evaluation Research on Different Data Distribution Patterns,” *Adv. Mater. Res.*, vol. 694–697, pp. 2794–2800, 2013, doi: 10.4028/www.scientific.net/amr.694-697.2794.
- [10] Q. Chen, “Application of K-Means Algorithm in Marketing,” *Adv. Econ. Manag. Polit. Sci.*, vol. 71, no. 1, pp. 178–184, 2024, doi: 10.54254/2754-1169/71/20241485.
- [11] M. A. W. Saputra and S. Harini, “Java Island Health Profile Clustering Using K-Means Data Mining,” *Int. J. Inf. Commun. Technol.*, vol. 8, no. 1, pp. 1–9, 2022, doi: 10.21108/ijoict.v8i1.606.
- [12] C. Larasvaty, “Cluster Analysis of Covid-19 in Indonesia Using K-Means Method,” *J. Dinda Data Sci. Inf. Technol. Data Anal.*, vol. 3, no. 1, pp. 36–39, 2023, doi: 10.20895/dinda.v3i1.822.
- [13] D. B. Dahl, J. Andros, and J. B. Carter, “Cluster Analysis via Random Partition Distributions,” 2021, doi: 10.48550/arxiv.2106.02760.
- [14] S. Developers, “Hierarchical Clustering Documentation.” 2023. [Online]. Available: <https://scikit-learn.org/stable/modules/clustering.html#hierarchical-clustering>
- [15] A. R. R. Alzahrani, “Impact of Dataset Scaling on Hierarchical Clustering: A Comparative Analysis of Distance-Based and Ratio-Based Methods,” *Int. J. Anal. Appl.*, vol. 22, p. 36, 2024, doi: 10.28924/2291-8639-22-2024-36.
- [16] N. Sunil Yadav, N. Dr. Munindra Kumar Singh, and N. Pankaj Kumar, “Data Mining Applications for Enhancing Healthcare Services: A Comprehensive Review,” *Int. J. Eng. Technol. Manag. Sci.*, vol. 7, no. 5, pp. 325–333, 2023, doi: 10.46647/ijetms.2023.v07i05.038.
- [17] A. Nowak-Brzezińska and I. Gaibei, “How the Outliers Influence the Quality of Clustering?,” *Entropy*, vol. 24, no. 7, p. 917, 2022, doi: 10.3390/e24070917.
- [18] S. Farisyi, M. Al Musadieq, H. N. Utami, and C. R. Damayanti, “Mapping Sustainability Reporting Disclosure Based on Determinants and Its Impact,” pp. 54–62, 2023, doi: 10.2991/978-94-6463-214-9_6.
- [19] K. Vahldiek and F. Klawonn, “Cluster-Centered Visualization Techniques for Fuzzy Clustering Results to Judge Single Clusters,” *Appl. Sci.*, vol. 14, no. 3, p. 1102, 2024, doi: 10.3390/app14031102.
- [20] H. N. Sofhya, “Cluster Analysis of Indonesian Provinces Based on Harvest Area and Rice Productivity Using Single Linkage Method,” *J. Sains Dan Teknol. Ind.*, vol. 20, no. 2, p. 549, 2023, doi: 10.24014/sitekin.v20i2.21737.
- [21] A. Muhsina and B. Joseph, “Evaluation of Different Clustering Techniques in Classifying the Vegetable Growing Panchayats of Ernakulam District, Kerala,” *Int. J. Plant Soil Sci.*, pp. 500–510, 2022, doi: 10.9734/ijpss/2022/v34i242666.
- [22] W.-C. Yang, J.-P. Lai, Y.-H. Liu, Y.-L. Lin, H.-P. Hou, and P.-F. Pai, “Using Medical Data and Clustering Techniques for a Smart Healthcare System,” *Electronics*, vol. 13, no. 1, p. 140, 2023, doi: 10.3390/electronics13010140.
- [23] M. Puli *et al.*, “Exploring Optimal Cluster Quality in Health Care Data (HCD): Comparative Analysis utilizing k-means Elbow and Silhouette Analysis,” *Int. J. Chem. Biochem. Sci.*, vol. 25, no. 16, pp. 48–60, 2024, [Online]. Available: www.iscientific.org/Journal.html
- [24] L. Zhang, O. J. Clay, S.-Y. Lee, and C. R. Howell, “Analyzing Multiple Social Determinants of Health Using Different Clustering Methods,” *Int. J. Environ. Res. Public Health*, vol. 21, no. 2, p. 145, 2024, doi: 10.3390/ijerph21020145.
- [25] S. N. S. Z. Shah and M. M. Rosli, “Clustering algorithms for analysing electronic medical record: A mapping study,” *IAES Int. J. Artif. Intell.*, vol. 12, no. 4, pp. 1784–1792, 2023, doi: 10.11591/ijai.v12.i4.pp1784-1792.