

Implementasi Data Mining Algoritma *Hierarchical Clustering* Untuk Klasterisasi Penyakit Menular Pada Manusia

Bella Amalia¹, Novhirtamely Kahar²,

1Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Nurdin Hamzah,
Jambi, Indonesia

e-mail: bellaamallia2301@gmail.com, novmely@gmail.com

Abstrak

Penyakit Penyakit menular masih menjadi tantangan serius karena penyebarannya yang cepat dan berdampak luas pada kesehatan masyarakat. Penelitian ini bertujuan mengklasifikasikan penyakit menular berdasarkan kesamaan gejala pasien menggunakan algoritma *Hierarchical Clustering* pada data mining. Penelitian dilakukan di Puskesmas Paal Lima Kota Jambi dengan data mencakup usia, jenis kelamin, lokasi, gejala, dan diagnosis penyakit menular. Prosedur meliputi pengumpulan dan pembersihan data, normalisasi, perhitungan jarak antar pasien menggunakan *Euclidean Distance*, serta pembentukan kluster dengan metode *average linkage*. Hasil penelitian menghasilkan tujuh kluster utama yang mewakili penyakit demam berdarah, TBC, kolera, kusta, pneumonia, muntaber, dan malaria. Kluster yang terbentuk selanjutnya dianalisis berdasarkan usia, jenis kelamin, dan wilayah untuk mengidentifikasi pola penyebaran penyakit. Evaluasi validasi dilakukan dengan membandingkan hasil klasterisasi pada perangkat lunak *Orange* dengan perhitungan manual di *Microsoft Excel*. Hasilnya menunjukkan tingkat kesesuaian 100%, di mana seluruh data pasien berhasil terkelompok sesuai diagnosis aslinya tanpa kesalahan. Hal ini menegaskan bahwa algoritma *Hierarchical Clustering* mampu menghasilkan kluster yang akurat dan konsisten. Dengan demikian, metode terbukti sangat efektif dalam mengelompokkan penyakit menular berdasarkan gejala dan karakteristik pasien. Pendekatan ini dapat membantu tenaga medis dalam diagnosis awal, pengendalian, serta perencanaan sumber daya kesehatan yang lebih optimal.

Kata Kunci: data mining, hierarchical clustering, klasifikasi, penyakit menular.

Abstract

Infectious diseases remain a critical challenge due to their rapid transmission and significant impact on public health. This study aims to classify infectious diseases based on patient symptom similarities using the *Hierarchical Clustering* algorithm in data mining. The research was conducted at Puskesmas Paal Lima, Jambi City, using patient data that included age, gender, location, symptoms, and confirmed diagnoses of infectious diseases. The process involved data collection and cleaning, normalization, distance calculation between patients using *Euclidean Distance*, and cluster formation with the *average linkage* method. The results generated seven main clusters representing dengue fever, tuberculosis, cholera, leprosy, pneumonia, acute gastroenteritis, and malaria. These clusters were further analyzed by age, gender, and location to identify patterns of disease distribution. Validation was carried out by comparing the clustering results from the *Orange* software with manual calculations in *Microsoft Excel*. The findings demonstrated 100% accuracy, with all patient data correctly grouped according to their original diagnoses without any misclassification. This confirms that the *Hierarchical Clustering* algorithm can produce highly accurate and consistent clustering results. Therefore, the *Hierarchical Clustering* method is proven to be highly effective in classifying infectious diseases based on patient symptoms and characteristics. This approach can assist healthcare professionals in early diagnosis, disease control, and better allocation of healthcare resources.

Keywords: classification, cluster validation, data mining, hierarchical clustering, infectious diseases.

1. Pendahuluan

Penyakit menular merupakan isu kesehatan global yang terus mendapat perhatian karena peningkatan kasus akibat urbanisasi, perubahan iklim, dan mobilitas manusia. Penyakit ini disebabkan oleh agen biologis seperti virus, bakteri, atau parasit, dan dapat menular melalui

kontak fisik, udara, cairan tubuh, makanan, minuman, maupun gigitan hewan, dengan dampak luas terhadap kesehatan, sosial, dan ekonomi [1].

Seiring dengan kemajuan teknologi informasi, data mining membantu analisis data kesehatan untuk menemukan pola tersembunyi. Melalui Hierarchical Clustering, data dapat dikelompokkan berdasarkan kesamaan karakteristik guna mengidentifikasi risiko tinggi dan menyusun intervensi yang tepat.

Beberapa penelitian sebelumnya, seperti penelitian yang dilakukan oleh Muhammad Hafiz dan Sylvia Anggraini menggunakan variabel Provinsi, Malaria, TB Paru, Pneumonia, Kusta, Tetanus, Campak, Diare, dan DBD. Hasilnya, K-Means Clustering berhasil mengelompokkan provinsi dengan pola penyebaran penyakit serupa, sehingga bermanfaat untuk memahami faktor-faktor yang memengaruhi prevalensi penyakit di tingkat regional dan sebagai dasar perumusan kebijakan kesehatan. Kekurangannya, metode ini peka terhadap inisialisasi pusat kluster sehingga penggunaannya harus disesuaikan dengan kebutuhan dan karakteristik data [2]. Dan penelitian yang dilakukan oleh Indira Nur Amalia, Yuyun Umaidah, dan Rini Mayasari menggunakan variabel Nama Kabupaten, Nama Kecamatan, Bentuk Penyakit (HIV/AIDS, IMS, DBD, Diare, TB, Malaria), Jumlah Korban, Satuan, dan Tahun. Hasil klasterisasi menghasilkan 3 kluster: klaster 1 dengan 6 kecamatan kategori sangat rawan, klaster 2 dengan 15 kecamatan kategori rawan, dan klaster 3 dengan 9 kecamatan kategori tidak rawan. Kekurangannya, penelitian ini belum menggunakan metode atau tools pembandingan serta cakupan data masih terbatas [3]. Dan penelitian yang dilakukan oleh Maria Tesalonika Siregar, Gusti Made Arya Sasmita, dan Gusti Agung Ayu Putri meneliti penyakit Polio, Tuberkulosis, dan HIV. Hasilnya menunjukkan metode terbaik adalah regresi linier sederhana dan *support vector regression*, dengan akurasi MAPE <10% pada polio dan tuberkulosis, serta akurasi MAPE 20–50% pada HIV [4].

2. Metode Penelitian

2.1. Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif karena bersifat objektif dan sistematis, memungkinkan analisis numerik yang dapat diuji secara statistik. Metode ini sesuai dengan penggunaan algoritma Hierarchical Clustering untuk mengelompokkan penyakit menular, karena mampu mengolah data besar dan menilai kualitas pengelompokan secara matematis.

Penelitian ini merupakan studi kasus di Puskesmas yang menganalisis data penyakit menular berdasarkan umur, jenis kelamin, gejala, dan diagnosis. Dengan algoritma Hierarchical Clustering, penelitian ini bertujuan mengidentifikasi kelompok penyakit dengan gejala serupa guna mendukung strategi pencegahan dan penanganan yang lebih efektif.

2.2 Tahap Penelitian

Tahap penelitian merupakan gambaran sistematis dari tahapan penelitian yang dilakukan untuk mencapai tujuan yang telah ditetapkan. Dalam penelitian “*Implementasi Data Mining Algoritma Hierarchical Clustering Untuk Klasifikasi Penyakit Menular Pada Manusia*”, kerangka kerja mencakup proses dari pengumpulan data hingga analisis hasil klasterisasi. Tahapan yang akan dilakukan dalam penelitian ini di deskripsikan



Gambar 2.1 Kerangka Kerja Penelitian

1. Tahap Pengumpulan Data

Tahap ini Mengumpulkan dataset pasien yang mengandung informasi gejala dan diagnosis penyakit menular. Data diperoleh melalui rekam medis serta hasil wawancara langsung dengan salah satu tenaga medis, yaitu dr. Fatan Chammah, AM.Kep, guna memastikan keakuratan dan relevansi data yang digunakan dalam penelitian ini.

2. Studi Literatur

Tahap Studi literatur merupakan tahap dalam penelitian yang bertujuan untuk meninjau penelitian sebelumnya yang relevan dengan topik yang sedang dikaji.

3. Tahap Preprocessing Data

Tahap ini membersihkan data dari nilai kosong atau duplikat, melakukan transformasi data agar dapat digunakan dalam algoritma *Hierarchical Clustering*.

4. Tahap Penerapan Algoritma *Hierarchical Clustering*

Tahap ini menentukan metode pengukuran jarak atau class target, menggunakan metode *Hierarchical Clustering* dan membentuk *grafik* untuk memahami hubungan antar klaster.

5. Tahap Analisis Hasil Klasterisasi

Pada tahap ini mengidentifikasi pola klaster yang terbentuk dan menganalisis karakteristik tiap klaster berdasarkan gejala penyakit.

6. Tahap Evaluasi

Pada tahap ini memvalidasi hasil dengan membandingkan dengan perhitungan manual *Microsoft Excel* atau dengan informasi yang ada.

7. Tahap Laporan

Tahap ini menyusun hasil penelitian dalam bentuk laporan dan visualisasi *grafik* dan memberikan interpretasi tentang pola penyakit menular yang ditemukan.

2.3. Data Dan Sumber Data

Dalam kegiatan penelitian, teknik pengumpulan data yang dipakai dalam memenuhi kebutuhan data atau informasi pada penelitian ini sebagai berikut:

Pengumpulan data sekunder yang dilakukan dengan mengambil dataset yang telah tersedia dari sumber terpercaya seperti Puskesmas yang mengidap penyakit menular, termasuk gejala yang dialami dan faktor risikonya.

1. Pengumpulan data primer yaitu jika dataset sekunder tidak tersedia atau kurang memenuhi kebutuhan penelitian, maka data dapat dikumpulkan langsung dari sumbernya melalui metode seperti, Surve atau kuesioner dan wawancara dengan ahli kesehatan.
2. Pengolahan dan Penyimpanan Data setelah data dikumpulkan, baik sekunder maupun primer data akan disimpan dalam bentuk format CSV, *Excel* atau database yang dapat diimpor ke dalam perangkat lunak seperti *Orange*.

2.4. Hierarchical Clustering

Hierarchical Clustering adalah metode pengelompokan data yang bekerja dengan menggabungkan objek-objek yang memiliki tingkat kemiripan tertinggi. Proses ini berlangsung secara bertahap hingga terbentuk struktur menyerupai pohon (tree) dengan hirarki dari objek paling mirip hingga paling tidak mirip. [5].

Hierarchical Clustering adalah metode pengelompokan data yang dimulai dengan menggabungkan dua atau lebih objek yang memiliki tingkat kemiripan tertinggi. [6].

Hierarchical clustering merupakan salah satu metode clustering yang didasarkan pada struktur seperti dendrogram, yaitu membagi atau menggabungkan data seperti cabang pohon dengan mengelompokkan dua atau lebih data yang memiliki kesamaan terdekat [7].

2.4.1. Average Linkage

Metode *average linkage* menghitung jarak antara dua cluster dengan cara mengambil rata-rata jarak antara setiap anggota dalam satu cluster terhadap semua anggota cluster lainnya [8]. *Average linkage* mengambil kesamaan antar cluster berdasarkan nilai rata-rata kesamaan. [9]. Metode Average Linkage merupakan variasi dari Single Linkage dan Complete Linkage yang menghitung jarak antar klaster berdasarkan rata-rata jarak antar semua pasangan objek [10]. Metode *average linkage* merupakan bagian dari metode Hierarchical clustering dimana pengelompokkannya dengan mencari rata rata dari jarak diantara pengamatan suatu objek [11].

2.5. Clustering

Clustering adalah metode dalam data mining yang bertujuan mengelompokkan data berdasarkan kemiripan karakteristik antar objek [12]. *Clustering* adalah pengelompokan data yang memiliki kemiripan antara satu data dengan data lainnya ke dalam klaster atau kelompok [13].

Clustering (klasterisasi) adalah proses mengelompokkan objek berdasarkan data yang menunjukkan hubungan antar objek, dengan tujuan memaksimalkan kesamaan dalam satu klaster [14]. *Clustering* bertujuan mengelompokkan data dengan karakteristik serupa ke dalam satu klaster dan memisahkan data yang berbeda karakteristik [15].

2.6. Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data [16]. Metode klasifikasi dapat digunakan untuk membuat prediksi menggunakan data terpadat dari Igracias [17]. Klasifikasi merupakan teknik supervised dalam data mining [18].

Metode Klasifikasi adalah metode data mining untuk memprediksi kelas objek berdasarkan atribut tertentu. [19]. merupakan suatu teknik untuk pengelompokan data menjadi beberapa bagian ataupun kelas [20].

3. Hasil dan Pembahasan

3.1. Perhitungan Manual

a. Normalisasi Dataset Menjadi Binner

Tabel dibawah ini menyajikan data gejala penyakit menular yang telah dinormalisasi menggunakan simbol (G1, G2, dan seterusnya) guna mempermudah proses pengolahan data dalam sistem klasifikasi penyakit.

Table 3.1 Variabel

Gejala	Simbol
Demam tinggi mendadak Nyeri otot, sendi, dan belakang mata Muncul bintik merah di kulit dan Mimisan atau gusi berdarah.	G1
Demam Tinggi hingga 40,55 derajat celcius, batuk kering atau berdahak, nyeri dada dan Sesak napas	G2
Batuk lebih dari 2 minggu, Batuk berdahak atau berdarah	G3
Diare, Mual Dan Muntah, Sakit Perut, Buang air besar disertai darah	G4

Gejala	Simbol
Hipoglikemia (gula darah rendah), Elektrolit dalam tubuh tidak seimbang, Dehidrasi dan kelelahan	G5
Demam tinggi berulang, Menggigil dan berkeringat	G6
Batuk lebih dari 2 minggu, Batuk berdahak atau berdarah	G7

Tabel berikut menampilkan data pasien yang telah dinormalisasi berdasarkan kemunculan gejala, dengan representasi biner (0 dan 1) sesuai simbol G1 hingga G7 untuk setiap pasien.

Table 3.2 Dataset Normalisasi

No	Nama	G1	G2	G3	G4	G5	G6	G7
1	P1	1	0	0	0	0	0	0
2	P2	0	1	0	0	0	0	0
3	P3	0	0	1	0	0	0	0
4	P4	0	0	1	0	0	0	0
5	P5	0	0	1	0	0	0	0
...	...	...	...	...	...	...	...	...
98	P98	0	0	0	0	1	0	0
99	P99	0	0	0	0	0	0	1
100	P100	0	1	0	0	0	0	0

b. Mencari Euclidean Distance

Pada langkah ini akan ditentukan jarak antar pasien menggunakan rumus

$$d = \sqrt{(x_1 - x_2)^2 + (y_2 - y_1)^2}$$

Untuk *distance* akan dibentuk menjadi matrik 100x100, karna datanya ada 100. Setiap pasien menghitung jarak antar pasien lain dan jarak ke diri sendiri, sebagai berikut :

$$P1 - P1 = \sqrt{(1 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2} = 0$$

$$P1 - P2 = \sqrt{(1 - 0)^2 + (0 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2} = 1,4$$

$$P1 - P3 = \sqrt{(1 - 0)^2 + (0 - 0)^2 + (0 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2} = 1,4$$

$$P1 - P27 = \sqrt{(1 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2} = 0$$

$$P1 - P100 = \sqrt{(1 - 0)^2 + (0 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2} = 1,4$$

c. Mencari Cluster Dengan Average Linkage

Proses pencarian kluster dilakukan dengan metode *average linkage*, yaitu menghitung rata-rata jarak antar pasien. Metode ini dipilih karena menggunakan rata-rata dari seluruh jarak antar pasangan titik pada dua kluster sehingga hasil pengelompokan lebih seimbang.

Table 3.3 Mencari Cluster Awal

Jarak	Target	Average
P1	P2	$(0+1,4)/2 = 0,7$
P1	P3	$(0+1,4)/2 = 0,7$
P1	P4	$(0+1,4)/2 = 0,7$
P1	P5	$(0+1,4)/2 = 0,7$
....		
P1	P97	$(0 + 0)/2 = 0$
P1	P98	$(0+1,4)/2 = 0,7$
P1	P99	$(0+1,4)/2 = 0,7$
P1	P100	$(0+1,4)/2 = 0,7$

d. Mencari Pasangan Cluster

Tabel berikut menunjukkan perhitungan dalam menentukan pasangan kluster, di mana sebelumnya telah ditentukan titik awal kluster menggunakan rumus *average linkage*. Selanjutnya, pasangan kluster ditentukan kembali dengan rumus yang sama, yaitu menghitung rata-rata jarak antar pasien.

Table 3.3 Mencari Pasangan Cluster

Jarak	Target	Average
P1 - 23	P2	$(1,4 + 1,4)/2 = 1,4$
P1 - 23	P3	$(1,4 + 1,4)/2 = 1,4$
P1 - 23	P4	$(1,4 + 1,4)/2 = 1,4$
P1 - 23	P5	$(1,4 + 1,4)/2 = 1,4$
P1 - 23	P6	$(1,4 + 1,4)/2 = 1,4$

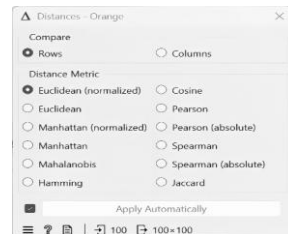
Jarak	Target	Average
P1 - 23	P97	$(0 + 0)/2 = 0$
P1 - 23	P98	$(1,4 + 1,4)/2 = 1,4$
P1 - 23	P99	$(1,4 + 1,4)/2 = 1,4$
P1 - 23	P100	$(1,4 + 1,4)/2 = 1,4$

3.2. Implementasi Orange



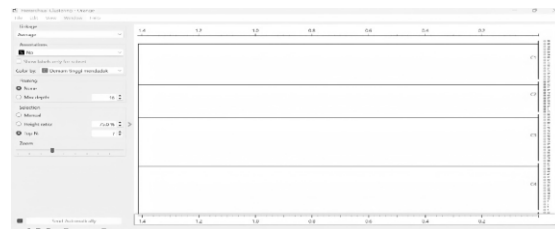
Gambar 3.1 Select Cell Data

Selanjutnya pada gambar diatas Lakukan pemilihan *cell* yang berupa data dari variabel yang akan di *clustering*.



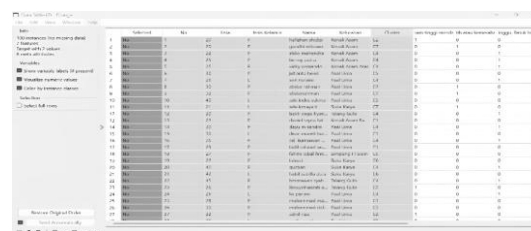
Gambar 3.2 Distance

Gambar diatas menunjukkan pengaturan *Distance* pada aplikasi Orange dengan opsi *Rows* untuk perhitungan jarak antar individu, serta metode *Euclidean (normalized)* sebagai matriks jarak yang mempertimbangkan normalisasi.



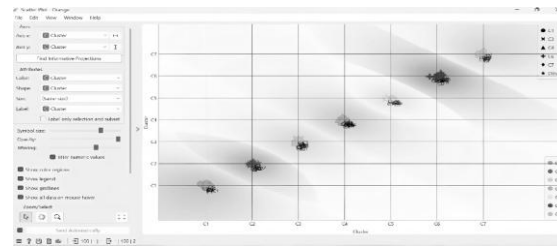
Gambar 3.3 Hierarchical Clustering

Gambar di atas menampilkan hasil *Hierarchical Clustering* dengan metode *Average Linkage* pada Orange. Dengan *max depth* 16 dan pemotongan 75% tinggi grafik, terbentuk tujuh kluster utama (C1–C7). Pewarnaan berfungsi sebagai pembeda visual, menunjukkan keberhasilan pengelompokan data berdasarkan kemiripan karakteristik.



Gambar 3.4 Hasil Cluster

Gambar di atas menunjukkan hasil *Hierarchical Clustering* menggunakan Orange, yang mengelompokkan pasien ke dalam tujuh kluster (C1–C7) berdasarkan kemiripan gejala (sakit tenggorokan, demam, batuk). Data biner (0 = tidak, 1 = ya) memperlihatkan pola kesamaan karakteristik pasien untuk analisis identifikasi penyakit lebih lanjut.



Gambar 3.5 Hasil Grafik

Gambar di atas menunjukkan hasil *clustering* pada aplikasi Orange, di mana data terbagi menjadi tujuh kluster (C1–C7) dengan warna berbeda. Titik mewakili individu berdasarkan kemiripan karakteristik, sementara area berwarna menandai batas antar kluster. Visualisasi ini memperjelas distribusi dan keterpisahan tiap kluster.

4. Kesimpulan

Berdasarkan hasil penelitian, bahwa algoritma Hierarchical Clustering mampu mengelompokkan pasien berdasarkan kemiripan gejala penyakit menular, sehingga pola hubungan antarpenyakit terlihat lebih jelas. Proses klusterisasi melalui aplikasi Orange menghasilkan kelompok yang terstruktur dan mudah dipahami. Metode ini efektif untuk klasifikasi penyakit seperti TBC, Kusta, Malaria, Kolera, Pneumonia, Demam Berdarah, dan Muntaber, menggunakan data dari Puskesmas Paal Lima Kota Jambi. Keunggulannya, metode ini tidak memerlukan jumlah kluster awal, sehingga hasil lebih sesuai dengan data dan dapat mendukung strategi pencegahan serta pengendalian penyakit menular.

5. Saran

Berdasarkan hasil penelitian dan implementasi *Hierarchical Clustering*, terdapat beberapa poin penting: perangkat dan aplikasi perlu dirawat serta dibackup rutin agar sistem tetap stabil dan aman. Sistem ini dapat dikembangkan menjadi aplikasi bagi tenaga medis atau staf puskesmas. Pengumpulan data gejala dan diagnosis perlu diperluas untuk meningkatkan akurasi analisis. Kekurangan sistem perlu dicatat untuk perbaikan. Penelitian selanjutnya disarankan membandingkan metode lain seperti *K-Means*, *DBSCAN*, *Decision Tree*, atau *Naïve Bayes* guna meningkatkan akurasi klasifikasi penyakit menular.

Daftar Pustaka

- [1] D. Pratiwi *et al.*, “Penyakit Menular Menggunakan Fuzzy C Means,” vol. 8, no. 4, pp. 7998–8005, 2024.
- [2] S. A. Muhammad Hafiz, “KLASTERISASI PENYAKIT MENULAR DI INDONESIA MENGGUNAKAN METODE K-MEANS CLUSTERING,” *J-Com (Journal Comput.*, vol. 15, no. 1, pp. 50–57, 2024.
- [3] I. Nur Amalia, Y. Umaidah, and R. Mayasari, “Penerapan Data Mining Untuk Klusterisasi Daerah Rawan Penyakit Menular Di Kabupaten Karawang Dengan Menggunakan Algoritma K-Means,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 4, pp. 5582–5591, 2024, doi: 10.36040/jati.v8i4.9953.
- [4] M. T. Siregar, G. Made Arya Sasmita, and G. Agung Ayu Putri, “Perbandingan Analisis Metode Peramalan Jumlah Kasus Penyakit Menular di Kota Bandung (Studi Kasus: Dinas Kesehatan Kota Bandung),” *J. Ilm. Teknol. dan Komput.*, vol. 3, no. 1, p. 6, 2022.
- [5] Anindya Khrisna Wardhani, “Implementasi Algoritma K-Means Untuk Pengelompokkan Penyakit Pasien Pada Puskesmas Kajen Pekalongan,” *J. Transform.*, vol. 14, pp. 30–37, 2016.

- [6] N. K. Zuhail, "Study Comparison K-Means Clustering Dengan Algoritma Hierarchical Clustering," *Pros. Semin. Nas. Teknol. dan Sains*, vol. 1, pp. 200–205, 2022.
- [7] B. M. Kumarahadi, H. Pratiwi, and S. Subanti, "Penerapan Metode Hierarchical Clustering Untuk Pengelompokan Kota/Kabupaten di Indonesia Berdasarkan Indikator Kemiskinan," *J. Teknol. Inf. dan Komun.*, vol. 11, no. 2, p. 13, 2023, doi: 10.30646/tikomsin.v11i2.754.
- [8] K. D. Larasati, "Analisis Kluster Hirarki dengan Metode Average Linkage," *Medium*, vol. 13, no. 6, pp. 849–856, 2019, [Online]. Available: <https://medium.com/@dhea.larasati326/analisis-kluster-hirarki-dengan-metode-average-linkage-a6add8889594>
- [9] E. Novia and P. Hendikawati, "Average Linkage Hierarchical Cluster untuk Pengelompokan Kabupaten / Kota di Provinsi Jawa Tengah Berdasarkan Variabel Pencegahan Terjadinya Stunting," *Prism. Pros. Semin. Nas. Mat.*, vol. 7, pp. 702–711, 2024, [Online]. Available: <https://proceeding.unnes.ac.id/prisma/article/view/3021/2484>
- [10] Musdalifah M. Ramly, S. Sudarmin, and B. Poerwanto, "Analisis Hierarchical Clustering Multiscale Bootstrap (Kasus: Indikator Kemiskinan di Provinsi Sulawesi Selatan Tahun 2020)," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 4, no. 3, pp. 142–152, 2022, doi: 10.35580/variensiunm26.
- [11] R. P. Butar Butar, "Analisis Hierarchical Dan Non-Hierarchical Clustering Untuk Pengelompokan Potensi Ekonomi Kelautan Indonesia 2021," *J. Sist. dan Teknol. Inf.*, vol. 11, no. 3, p. 543, 2023, doi: 10.26418/justin.v11i3.67283.
- [12] U. Ma'rifatin, "Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien Pada Puskesmas Warujayeng," *Semin. Nas. Inov. Teknol.*, no. 2549–7952, pp. 285–291, 2020.
- [13] D. Ariyanto, "Data Mining Menggunakan Algoritma K-Means untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut," *J. Sistim Inf. dan Teknol.*, vol. 4, pp. 13–18, 2022, doi: 10.37034/jsisfotek.v4i1.117.
- [14] F. B. Lazuardi and Hanna Prillysca Chernovita, "Analisis persebaran dan visualisasi penyakit Infeksi Saluran Pernapasan Akut (ISPA) dengan metode K-Means Clustering pada Provinsi Jawa Tengah," vol. 04, no. 52, pp. 161–172, 2025, doi: 10.24246/itexplore.v4i2.2025.pp161-172.
- [15] M. Hafiz Ziqrullah, Andri, S. Dian Purnamasari, and I. Zuhri Yadi, "Clustering Data Penyakit Pasien Pada Puskesmas Mulyaguna Menggunakan Algoritma K-Means," vol. 6, no. 2, pp. 608–616, 2024, [Online]. Available: <http://ejournal.sisfokomtek.org/index.php/jumin>
- [16] Y. Arvita, Suyanti, and A. Siswanto, "Classification of Monkeypox Disease Using the K-Nearest Neighbor Algorithm," *Process. J. Ilm. Sist. Informasi, Teknol. Inf. dan Sist. Komput.*, vol. 19, no. 2, pp. 137–142, 2024.
- [17] B. I. Nugroho, Z. Ma'arif, and Z. Arif, "Tinjauan Pustaka Sistematis: Penerapan Data Mining Metode Klasifikasi Untuk Menganalisa Penyalahgunaan Sosial Media," *J. Sist. Inf. dan ...*, vol. 3, no. 2, pp. 46–51, 2022, [Online]. Available: <http://journal.peradaban.ac.id/index.php/jsitp/article/download/1265/860>
- [18] N. A. Ramadhani and H. A. Rosyid, "Algoritma Data Mining untuk Klasifikasi Data," *J. Inov. Teknol. dan Edukasi Tek.*, vol. 2, no. 12, pp. 550–556, 2022, doi: 10.17977/um068v2i122022p550-556.
- [19] A. A. Abidin, R. Goejantoro, and M. Fathurahman, "Klasifikasi Penyakit Tuberkulosis Menggunakan Metode Naive Bayes (Studi Kasus: Data Pasien Di Puskesmas Petung Kabupaten Penajam Paser Utara)," *Ekspansional*, vol. 14, no. 1, p. 11, 2023, doi: 10.30872/ekspansional.v14i1.1031.
- [20] A. N. P. Haryadi, W. Setiawan, and D. A. Fatah, "Penerapan Data Mining untuk Klasifikasi Penyakit Diabetes Menggunakan Metode Decision Tree," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 7, no. 6, pp. 1484–1495, 2024, doi: 10.32672/jnkti.v7i6.8145.