

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia menghadapi fenomena beban ganda penyakit (*double burden of disease*) yang semakin kompleks, dimana penyakit menular dan tidak menular secara simultan memberikan tekanan signifikan terhadap sistem kesehatan nasional. Data Survei Kesehatan Indonesia (SKI) 2023 yang dirilis oleh Kementerian Kesehatan mencatat prevalensi hipertensi berdasarkan pengukuran tensimeter pada penduduk usia ≥ 18 tahun sebesar 30,8%, dengan distribusi yang bervariasi antar kelompok usia 10,7% pada usia 18–24 tahun dan 17,4% pada usia 25–34 tahun (Kementerian Kesehatan Republik Indonesia, 2023). Lebih mengkhawatirkan, penyakit tidak menular (PTM) bertanggung jawab atas 75% kematian di Indonesia, dengan penyakit kardiovaskular seperti serangan jantung dan stroke merenggut hampir 800.000 nyawa setiap tahunnya. Pada tahun 2023, pembiayaan penyakit katastrofik menghabiskan Rp34,8 triliun dari dana Jaminan Kesehatan Nasional (JKN), dengan jantung dan stroke menyerap porsi terbesar sebesar Rp22,8 triliun (Litha, 2025). Di sisi lain, prevalensi stunting meskipun menunjukkan tren penurunan dari 21,5% pada 2023 menjadi 19,8% pada 2024, masih menjadi tantangan besar dengan target nasional 14,2% pada tahun 2029 (Indonesia, 2025).

Kompleksitas beban kesehatan ini menuntut transformasi dalam pengelolaan data kesehatan di tingkat fasilitas primer. Implementasi sistem digital seperti Electronic Medical Record (EMR) dan platform SatuSehat oleh Kementerian Kesehatan telah mengakselerasi pengumpulan data pasien di Puskesmas (Hidayat et al., 2024). Penelitian menunjukkan bahwa implementasi EMR secara signifikan meningkatkan kualitas informasi, kualitas sistem, dan efisiensi pelatihan tenaga kesehatan dalam pengelolaan data medis (Muchlis, 2024). Peraturan Menteri Kesehatan Nomor 24 Tahun 2022 Pasal 5 mewajibkan seluruh fasilitas kesehatan untuk menyelenggarakan pencatatan elektronik paling lambat 31 Desember 2023, menjadi landasan hukum berkembangnya EMR di Indonesia (Zaman, 2023). Namun demikian, tantangan implementasi masih dihadapi, terutama terkait kesiapan sumber daya manusia, infrastruktur jaringan yang tidak stabil, serta keterbatasan komputer di beberapa Puskesmas (Kementerian Kesehatan Republik Indonesia, 2024).

Seiring dengan meningkatnya volume data pasien yang terakumulasi melalui sistem EMR, kebutuhan akan analisis data yang mendalam dan berbasis bukti menjadi semakin mendesak. Metode data mining menawarkan solusi untuk mengekstraksi pola tersembunyi dari dataset besar yang tidak dapat diidentifikasi melalui pendekatan manual konvensional. Dalam konteks kesehatan, teknik *clustering* telah terbukti efektif untuk segmentasi pasien berdasarkan kemiripan karakteristik, memungkinkan personalisasi intervensi dan alokasi sumber daya yang lebih tepat sasaran (Williams, 2025). Penelitian oleh Moreira et al. (2018) mendemonstrasikan bahwa model hybrid data mining yang menggabungkan *clustering* dan klasifikasi dapat meningkatkan presisi dan sensitivitas dalam diagnosis penyakit dibandingkan dengan pendekatan tunggal (Moreira & Namen, 2018).

Algoritma Agglomerative Nesting (AGNES) sebagai metode hierarchical *clustering* memiliki keunggulan dalam membangun struktur hierarki *cluster* yang dapat divisualisasikan melalui dendrogram, memungkinkan identifikasi kelompok risiko dengan presisi lebih tinggi (Noble, 2024). AGNES efektif dalam menangkap struktur hierarkis data dan dapat menangani berbagai tipe data dengan berbagai ukuran jarak (Kurniawan, 2024). Penelitian Liu et al. (2021) mengembangkan metode AGNEP yang mengintegrasikan algoritma AGNES dengan Principal Component Analysis (PCA) untuk reduksi dimensi fenotip, menunjukkan superioritas dalam mendeteksi asosiasi genetik dibandingkan metode konvensional seperti MANOVA dan ANOVA (Liu et al., 2021). Dalam konteks fasilitas kesehatan, Alwani (2024) mengaplikasikan AGNES dengan complete linkage *clustering* untuk pemetaan fasilitas kesehatan di Jawa Barat, menghasilkan dua *cluster* optimal yang divalidasi dengan *silhouette coefficient* dan *Calinski-Harabasz Index* (Alwani, Fitrianto, et al., 2024).

Penelitian hierarchical *clustering* untuk segmentasi pasien di layanan kesehatan primer menunjukkan hasil yang menjanjikan. Studi oleh cluster analysis pada data kesehatan primer di Wales berhasil mengidentifikasi segmen populasi dengan kebutuhan kesehatan unik berdasarkan pola utilisasi layanan, memberikan intelijen yang berguna untuk perencanaan layanan kesehatan dan integrasi perawatan (Nnoaham & Cann, 2020). Untuk evaluasi kualitas *cluster*, kombinasi Silhouette Score dan Davies-Bouldin Index (DBI) telah menjadi standar dalam

literatur. Silhouette Score mengukur seberapa baik setiap objek ditempatkan dalam *clusternya* sendiri dibandingkan *cluster* lain, sementara DBI mengevaluasi rasio separasi antar *cluster* semakin mendekati 0, semakin baik kualitas *cluster* (Amrulloh, 2022).

Implementasi *data mining* di Puskesmas Indonesia telah menunjukkan hasil positif dalam berbagai konteks. Penelitian Pane et al. (2024) menerapkan K-Means *clustering* untuk memprediksi pengadaan obat di Puskesmas Mandala, yang berhasil mengelompokkan obat ke dalam kategori permintaan tinggi, sedang, dan rendah guna meningkatkan efisiensi stok logistik (Pane et al., 2024). Studi serupa di Puskesmas Jember menggunakan K-Means untuk *clusterisasi* berdasarkan indikator kesehatan ibu dan anak, memberikan referensi bagi Dinas Kesehatan (Prakoso et al., 2023). Pane et al. mengimplementasikan K-Means untuk memprediksi pengadaan obat di Puskesmas Mandala dengan hasil validitas yang baik (Pane et al., 2024).

Meskipun demikian, mayoritas penelitian *clustering* di Puskesmas Indonesia masih menggunakan K-Means yang memiliki keterbatasan dalam penentuan jumlah *cluster* secara *a priori* dan sensitivitas terhadap inisialisasi *centroid* (No et al., 2025). Penelitian yang memanfaatkan *hierarchical clustering*, khususnya AGNES, untuk segmentasi pasien individual di layanan kesehatan primer masih sangat terbatas.

Puskesmas Kamoning sebagai fasilitas kesehatan primer di Kabupaten Sampang memiliki karakteristik pasien yang beragam: lansia dengan komorbiditas ganda, pasien muda dengan kunjungan rutin, hingga kelompok risiko tinggi yang belum tersegmentasi dengan baik. Dengan rata-rata 150–200 kunjungan pasien per minggu dan proporsi kasus tidak menular di atas 40%, pengembangan model analitik berbasis data mining menjadi kebutuhan strategis untuk transformasi layanan dari pendekatan reaktif menjadi proaktif dan berbasis bukti.

Hasil penelitian diharapkan memberikan kontribusi ganda: secara akademis, mengisi kesenjangan literatur mengenai penerapan *hierarchical clustering* pada data pasien individual di Puskesmas Kamoning; secara praktis, menghasilkan model segmentasi pasien yang dapat diterjemahkan menjadi aksi operasional konkret seperti penentuan skema screening prioritas, penjadwalan edukasi kesehatan, optimalisasi pengadaan obat, dan penempatan SDM kesehatan berdasarkan kebutuhan *cluster*. Model ini selanjutnya dapat menjadi referensi untuk

pengembangan dashboard kesehatan digital, pengajuan anggaran berbasis bukti, dan advokasi kebijakan di tingkat regional maupun nasional.

1.2 Perumusan Masalah

Pada bagian perumusan masalah, diuraikan mengenai permasalahan, metode usulan serta pertanyaan penelitian yang diuraikan sebagai berikut :

1.2.1 Permasalahan

Berdasarkan uraian latar belakang, dapat diambil permasalahan penelitian di antaranya sebagai berikut :

1. Puskesmas memiliki ribuan data rekam medis berisi informasi penting seperti diagnosa, tekanan darah, hingga demografi pasien. Sayangnya, data ini seringkali hanya berakhir sebagai arsip laporan bulanan ke dinas. Cerita tentang pola penyakit yang sedang tren atau kelompok pasien mana yang paling rentan "terkubur" dalam tumpukan data tersebut karena tidak pernah dianalisis secara mendalam.
2. Dalam penentuan prioritas intervensi kesehatan, seperti program skrining atau pemantauan intensif, keputusan seringkali masih didasarkan pada pertimbangan subjektif dan rutinitas pelayanan. Belum adanya mekanisme pengelompokan (*clustering*) pasien secara objektif menyebabkan kesulitan dalam memetakan kelompok risiko secara presisi, yang berpotensi membuat alokasi sumber daya layanan menjadi kurang efisien dan tidak tepat sasaran.
3. Proses perencanaan program kesehatan dan logistik farmasi cenderung bersifat pengulangan, mengikuti pola lama tanpa menyesuaikan dengan kondisi data yang sebenarnya terjadi. Ketidadaan analisis yang mampu memperkirakan kebutuhan membuat manajemen Puskesmas lebih sering menangani masalah saat sudah kejadian, daripada bersiap sejak awal mengantisipasi kebutuhan masyarakat berdasarkan data.
4. Belum adanya pemodelan pengelompokan pasien berbasis data rekam medis di Puskesmas Kamoning yang mampu menunjukkan struktur kelompok pasien secara jelas, sehingga pola hubungan antar kelompok risiko belum terlihat

dan sulit dijadikan dasar pengambilan keputusan yang terarah.

1.2.2 Metode Usulan

Untuk mengatasi permasalahan pengelolaan data serta belum optimalnya pemanfaatan informasi mengenai pola penyakit dan kelompok pasien, penelitian ini mengusulkan penerapan metode Hierarchical *Clustering* dengan algoritma Agglomerative Nesting (AGNES) pada data rekam medis pasien di Puskesmas Kamoning. Penelitian ini berfokus pada analisis data rekam medis pasien tahun 2024 dengan menggunakan sembilan variabel penelitian, yaitu umur, jenis kelamin, tekanan darah, berat badan, tinggi badan, respiratory rate, heart rate, diagnosa, dan poli. Dalam proses pembentukan cluster, digunakan tujuh variabel numerik, yaitu umur, berat badan, tinggi badan, tekanan darah sistolik, tekanan darah diastolik, respiratory rate, dan heart rate. Sementara itu, variabel jenis kelamin, diagnosa, dan poli digunakan untuk membantu analisis karakteristik pada setiap cluster yang terbentuk. Metode AGNES digunakan untuk mengelompokkan pasien berdasarkan kemiripan kondisi medisnya, kemudian divisualisasikan dalam bentuk struktur hierarki atau dendrogram agar hubungan antar kelompok pasien dapat terlihat dengan lebih jelas. Hasil clusterisasi ini diharapkan menghasilkan segmentasi pasien yang lebih objektif dan mudah dipahami, sehingga dapat dimanfaatkan oleh Puskesmas Kamoning dalam menentukan prioritas skrining, memantau pasien penyakit kronis, menyusun program edukasi kesehatan, serta mengalokasikan sumber daya layanan secara lebih tepat sasaran dan berbasis data.

1.2.3 Pertanyaan Penelitian

Berdasarkan latar belakang dan permasalahan yang telah diuraikan, pertanyaan peneliti yang menjadi fokus dalam studi ini adalah sebagai berikut :

1. Bagaimana penerapan algoritma Hierarchical *Clustering* (AGNES) pada data rekam medis pasien tahun 2024 di Puskesmas Kamoning untuk mengelompokkan pasien berdasarkan karakteristik klinis?
2. Berapakah jumlah *cluster* optimal yang terbentuk pada data rekam medis pasien Puskesmas Kamoning menggunakan algoritma AGNES berdasarkan indeks validasi *cluster* yang digunakan?
3. Apa sajakah karakteristik profil kesehatan dari setiap *cluster* pasien yang terbentuk, dan rekomendasi layanan kesehatan apa yang dapat dirumuskan

untuk mendukung perencanaan program di Puskesmas Kamoning?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

Berdasarkan rumusan masalah yang ada, dapat ditentukan tujuan dari penelitian ini diantaranya adalah sebagai berikut :

1. Menerapkan algoritma *Hierarchical Clustering* (AGNES) pada data rekam medis pasien tahun 2024 di Puskesmas Kamoning untuk mengelompokkan pasien berdasarkan karakteristik klinis yang digunakan dalam penelitian.
2. Menentukan jumlah *cluster* optimal yang terbentuk pada data rekam medis pasien Puskesmas Kamoning menggunakan algoritma AGNES berdasarkan nilai indeks validasi *cluster* yang digunakan (misalnya Silhouette Coefficient dan Davies-Bouldin Index).
3. Menganalisis karakteristik profil kesehatan dari setiap *cluster* pasien yang terbentuk dan menyusun rekomendasi layanan kesehatan yang dapat dimanfaatkan untuk mendukung perencanaan program di Puskesmas Kamoning.

1.3.2 Manfaat

Berdasarkan rumusan masalah dan tujuan yang ada, dapat ditentukan manfaat dari penelitian ini diantaranya adalah sebagai berikut :

1. Bagi Puskesmas Kamoning, penelitian ini memberikan gambaran kelompok pasien yang lebih terstruktur beserta karakteristik utamanya sehingga dapat dimanfaatkan untuk menyusun prioritas skrining, merancang materi edukasi kesehatan yang lebih tepat sasaran, serta merencanakan kebutuhan layanan dan penempatan tenaga kesehatan sesuai komposisi *cluster* pasien di wilayah kerja puskesmas.
2. Bagi masyarakat, penelitian ini bermanfaat dalam tersedianya layanan yang lebih terarah dan proaktif, karena puskesmas dapat lebih cepat mengenali kelompok masyarakat berisiko dan merancang intervensi pencegahan maupun pemantauan rutin berdasarkan profil *cluster* yang ditemukan.
3. Bagi ilmu pengetahuan, penelitian ini menambah kontribusi empiris mengenai penerapan metode *Hierarchical Clustering*, khususnya algoritma AGNES, pada data rekam medis pasien di puskesmas,

sehingga dapat menjadi rujukan praktis dan titik awal bagi peneliti berikutnya yang ingin mengembangkan studi segmentasi pasien dan pemanfaatan data mining di bidang kesehatan.

1.4 Batasan Masalah

Berikut batasan-batasan yang akan diberikan agar permasalahan tidak terlalu luas dan tidak menyimpang sehingga sesuai dengan tujuan penulis berikut ini :

1. Penelitian ini menggunakan data rekam medis pasien dari Puskesmas Kamoning khusus periode tahun 2024.
2. Variabel yang digunakan dalam penelitian dibatasi pada sembilan atribut, yaitu umur, jenis kelamin, tekanan darah (sistolik dan diastolik), berat badan, tinggi badan, *respiratory rate*, *heart rate*, diagnosa, dan poli. Dalam proses pembentukan cluster, hanya tujuh variabel numerik yang digunakan, yaitu umur, tekanan darah sistolik, tekanan darah diastolik, berat badan, tinggi badan, *respiratory rate*, dan *heart rate*. Sementara itu, variabel jenis kelamin, diagnosa, dan poli digunakan untuk analisis karakteristik cluster.
3. Penelitian ini hanya menerapkan metode Hierarchical *Clustering* dengan algoritma Agglomerative Nesting (AGNES) untuk melakukan pengelompokan pasien berdasarkan kemiripan karakteristik klinis.
4. Evaluasi hasil pengelompokan difokuskan pada pengukuran kualitas *cluster* menggunakan indeks validasi *cluster* (misalnya Silhouette Coefficient dan Davies-Bouldin Index), serta analisis deskriptif terhadap profil kesehatan setiap *cluster* yang terbentuk.
5. Hasil penelitian berupa model komputasi untuk segmentasi pasien dan rekomendasi manajerial bagi perencanaan layanan kesehatan masyarakat di Puskesmas Kamoning, bukan sistem diagnosis klinis untuk penegakan diagnosis ataupun penentuan terapi pada pasien perorangan.

1.5 Sistematika Penulisan

Sistematika penulisan skripsi ini penulis susun dengan mengikuti aturan penulisan skripsi pada program studi S1 Sistem Informasi Universitas Trunodjoyo Madura. Berikut merupakan metodologi penulisan skripsi pada penelitian ini.

BAB I PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, metode usulan, pertanyaan penelitian, tujuan dan manfaat, batasan masalah, serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini menguraikan teori terkait Puskesmas, EMR, KDD, EDA, data mining dan *clustering*, metode *Hierarchical Clustering* (AGNES), evaluasi *clustering*, serta penelitian terdahulu.

BAB III METODE USULAN

Bab ini menjelaskan tahapan penelitian meliputi pengumpulan data, preprocessing, implementasi AGNES, penentuan jumlah *cluster*, evaluasi, dan analisis hasil.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil pengolahan data, proses *clustering*, evaluasi, serta analisis dan pembahasan hasil penelitian.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan penelitian dan saran untuk pengembangan selanjutnya.

DAFTAR PUSTAKA

Bab berisi tentang keterangan mengenai sumber rujukan yang digunakan dalam penyusunan Tugas Akhir. Pada penelitian ini daftar pustaka disusun secara otomatis dengan menggunakan aplikasi mendeley.

DAFTAR LAMPIRAN

Bagian ini mencakup lampiran yang berisi data dan informasi pendukung yang relevan dengan penelitian.