Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 6, No. 3, Maret 2026, Hal. 471-479
DOI: https://doi.org/10.52436/1.jpti.1477
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Pengelompokan Level Hipertensi Berbasis Tekanan Darah Menggunakan Algoritma K-Means Clustering

Egga Asoka*[1], Fathoni[2], Hadipurnawan Satria[3], Edith Anggina[4]

[1]Program Doktoral Ilmu Teknik, Universitas Sriwijaya; Manajemen Informatika,Politeknik Negeri Sriwijaya
[2,3]Fakultas Ilmu Komputer, Universitas Sriwijaya
[4]Siloam Hospital Lippo Village, Tangerang
Email: [1]03013622530005@student.unsri.ac.id, [2]fathoni@unsri.ac.id,
[3]hadipurnawan.satria@unsri.ac.id, [4]edith.anggina@gmail.com

Abstrak

Hipertensi merupakan salah satu komponen utama sindrom metabolik yang berkontribusi signifikan terhadap peningkatan risiko penyakit kardiovaskular. Deteksi dini dan pemetaan tingkat hipertensi menjadi penting untuk mendukung intervensi medis yang tepat. Penelitian ini bertujuan untuk menerapkan algoritma unsupervised learning Algoritma K-Means Clustering dalam mengelompokkan individu berdasarkan parameter tekanan darah sistolik dan diastolik. Dataset yang digunakan terdiri dari 1.878 catatan pasien, yang setelah proses pembersihan data menghasilkan 1.575 data unik. Data distandarisasi menggunakan StandardScaler, dan jumlah klaster optimal ditentukan melalui metode Elbow. Hasil klasterisasi menunjukkan empat klaster utama yang merepresentasikan segmentasi alami tekanan darah, mulai dari tekanan darah rendah hingga tinggi. Visualisasi dua dimensi dan reduksi dimensi menggunakan Principal Component Analysis (PCA) memperlihatkan pemisahan klaster yang relatif jelas. Temuan ini menunjukkan bahwa K-Means mampu mengidentifikasi struktur laten data tekanan darah secara objektif dan berpotensi menjadi dasar pengembangan sistem pendukung keputusan medis berbasis data untuk stratifikasi risiko hipertensi.

Kata kunci: *Clustering, Hipertensi, K-Means, Pembelajaran Mesin, Tekanan Darah*

Hypertension Level Grouping Based on Blood Pressure Using the K-Means Clustering Algorithm

Abstract

Hypertension is a major component of metabolic syndrome that significantly contributes to an increased risk of cardiovascular disease. Early detection and mapping of hypertension levels are crucial to support appropriate medical interventions. This study aims to apply the unsupervised learning algorithm K-Means Clustering to group individuals based on systolic and diastolic blood pressure parameters. The dataset used consists of 1,878 patient records, which, after data cleaning, results in 1,575 unique data points. The data were standardized using StandardScaler, and the optimal number of clusters was determined using the Elbow method. The clustering results show four main clusters representing the natural segmentation of blood pressure, ranging from low to high blood pressure. Two-dimensional visualization and dimensionality reduction using Principal Component Analysis (PCA) show relatively clear cluster separation. These findings indicate that K-Means is able to objectively identify the latent structure of blood pressure data and has the potential to be the basis for developing a data-driven medical decision support system for hypertension risk stratification.

Keywords: *Blood Pressure, Clustering, Hypertension, K-Means, Machine Learning*

1. PENDAHULUAN

Sindrom metabolik merupakan kumpulan kondisi yang saling berkaitan, seperti hipertensi, hiperglikemia, dislipidemia, dan obesitas sentral, yang secara signifikan meningkatkan risiko penyakit kardiovaskular, diabetes tipe 2, dan gangguan metabolik lainnya [1],[2]. Gejala sindrom ini umumnya tidak spesifik dan berkembang secara perlahan, sehingga seringkali tidak terdeteksi pada tahap awal. Oleh karena itu, deteksi dini menjadi aspek penting dalam upaya pencegahan dan manajemen penyakit yang efektif. Salah satu komponen utama sindrom metabolik adalah hipertensi, yakni kondisi tekanan darah yang meningkat secara persisten dan berpotensi menyebabkan

p-ISSN: 2775-4227
e-ISSN: 2775-4219

kerusakan organ vital seperti jantung, ginjal, dan otak [3],[4]. Hipertensi telah lama dikenal sebagai salah satu kontributor utama dalam penyakit jantung dan komplikasi vaskular. Berdasarkan panduan American College of Cardiology/American Heart Association (ACC/AHA), tekanan darah sistolik ≥140 mmHg dan/atau diastolik ≥90 mmHg dikategorikan sebagai hipertensi [3]. Beberapa faktor risiko utama meliputi gaya hidup tidak sehat, resistensi insulin, serta kelainan metabolik lainnya [5],[6],[7]. Berbagai penelitian telah menunjukkan adanya hubungan timbal balik antara hipertensi dan sindrom metabolik, termasuk keterlibatan resistensi insulin, disfungsi sistem saraf simpatis, dan peradangan kronis [8], [9].

Dengan berkembangnya teknologi kecerdasan buatan, metode pembelajaran mesin (machine learning) semakin banyak diterapkan dalam bidang kesehatan, khususnya untuk mendeteksi kondisi kronis berdasarkan data klinis. Penelitian ini berfokus pada implementasi algoritma unsupervised learning K-Means untuk mengelompokkan pasien berdasarkan tingkat tekanan darah serta faktor risiko metabolik lainnya, tanpa memerlukan label data awal. Pemanfaatan algoritma ini bertujuan untuk mengidentifikasi subpopulasi pasien yang memiliki kecenderungan risiko hipertensi berdasarkan kemiripan fisiologis dan parameter metabolik. Dalam penelitian ini digunakan dataset berisi 1.878 catatan pasien dengan parameter seperti tekanan darah sistolik dan diastolik, kadar glukosa puasa, kolesterol, dan data antropometri. Dengan penerapan algoritma K-Means, diharapkan dapat dibentuk klaster-klaster pasien dengan tingkat hipertensi yang berbeda, mulai dari tekanan darah normal hingga hipertensi berat, sebagai langkah awal menuju pengembangan sistem pendukung keputusan medis yang lebih cerdas dan adaptif. Kecerdasan buatan (Artificial Intelligence/AI), khususnya teknik machine learning, telah menunjukkan potensi besar dalam membantu diagnosis dan prediksi penyakit kronis. Salah satu algoritma yang sering digunakan adalah K-Means Clustering, yang dapat mempartisi data ke dalam beberapa kelompok berdasarkan kesamaan karakteristik [10],[11]. Prinsip dasar K-Means adalah meminimalkan variansi dalam klaster (intra-cluster variance) dan memaksimalkan perbedaan antar-klaster (inter-cluster distance) secara iteratif.

Studi oleh Ghosh dan Kumar[10], serta Ikotun et al.[11], menunjukkan bahwa K-Means efektif dalam mengelompokkan data medis yang kompleks dan tidak berlabel, khususnya dalam studi tentang risiko kardiometabolik. Dalam konteks hipertensi, K-Means telah digunakan untuk mengelompokkan pasien berdasarkan tekanan darah dan parameter metabolik lainnya guna membantu segmentasi risiko secara lebih objektif [12]. Sebagai perbandingan, berbagai penelitian lain telah mengandalkan pendekatan supervised learning, seperti Random Forest [13], Support Vector Machine (SVM) [14], dan XGBoost [15], untuk prediksi hipertensi. XGBoost, misalnya, dikenal karena kemampuannya dalam menangani hubungan nonlinier dan dataset besar [16], serta terbukti menghasilkan kinerja prediksi yang tinggi dalam berbagai studi klinis [15],[17]. Namun, banyak pendekatan supervisi tersebut memiliki keterbatasan dalam menangkap keragaman subpopulasi metabolik, terutama ketika tidak ada label kelas yang eksplisit. Oleh karena itu, pendekatan klasterisasi seperti K-Means menjadi alternatif penting dalam eksplorasi pola tersembunyi dalam data klinis. Penelitian ini mengambil posisi di antara pendekatan eksploratif dan prediktif dengan menekankan pentingnya segmentasi awal pasien sebagai landasan untuk sistem pendukung keputusan klinis berbasis data [18], [19].

Meskipun berbagai penelitian sebelumnya telah menerapkan pendekatan *supervised learning* seperti Random Forest, Support Vector Machine, dan XGBoost untuk prediksi hipertensi, pendekatan tersebut sangat bergantung pada ketersediaan label kelas yang akurat. Dalam praktik klinis, label hipertensi sering kali tidak konsisten atau belum tersedia pada tahap awal pemeriksaan. Penelitian yang secara khusus mengeksplorasi struktur alami data tekanan darah menggunakan pendekatan *unsupervised learning*, khususnya untuk segmentasi level hipertensi, masih relatif terbatas. Oleh karena itu, diperlukan pendekatan eksploratif berbasis data untuk mengidentifikasi pola laten tekanan darah tanpa ketergantungan pada label klinis.

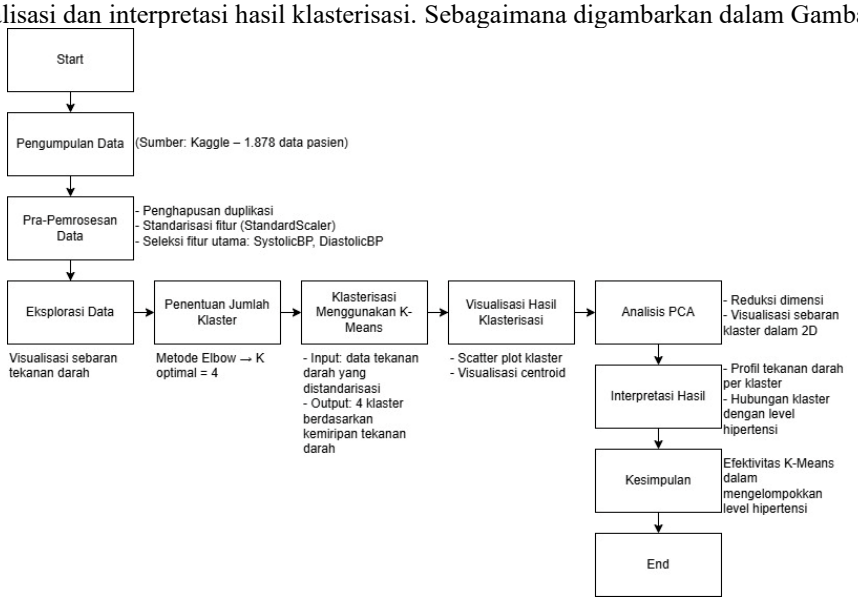
Berdasarkan tinjauan tersebut, penelitian ini bertujuan untuk mengelompokkan individu berdasarkan level tekanan darah sistolik dan diastolik menggunakan Algoritma K-Means Clustering, serta mengevaluasi kemampuan metode tersebut dalam membentuk segmentasi tekanan darah yang bermakna secara statistik dan klinis sebagai dasar pengembangan sistem pendukung keputusan medis berbasis data, dengan fokus eksplorasi struktur tekanan darah tanpa label masih relatif terbatas, khususnya dalam konteks populasi dengan karakteristik metabolik yang kompleks. Oleh karena itu, penelitian ini menawarkan kontribusi baru dengan menerapkan K-Means untuk membentuk klaster pasien berdasarkan tekanan darah sistolik dan diastolik secara data-driven, yang kemudian dapat menjadi fondasi bagi sistem klasifikasi atau intervensi medis yang lebih personal dan adaptif.

2. METODE PENELITIAN

Bagian ini menguraikan secara sistematis tahapan metodologi yang digunakan dalam penelitian ini, yang bertujuan untuk mengidentifikasi level hipertensi melalui pendekatan unsupervised learning menggunakan algoritma K-Means. Penelitian ini menggunakan pendekatan kuantitatif dengan desain potong lintang, dengan fokus pada pengelompokan pasien berdasarkan tekanan darah sistolik dan diastolik. Rangkaian metodologis

p-ISSN: 2775-4227
e-ISSN: 2775-4219

mencakup tahapan pemrosesan data, transformasi fitur, pelaksanaan algoritma K-Means, penentuan jumlah klaster optimal, serta visualisasi dan interpretasi hasil klasterisasi. Sebagaimana digambarkan dalam Gambar 1.



Gambar 1. Alur Penelitian

2.1. Pemrosesan Data

Dataset yang digunakan dalam penelitian ini diperoleh dari repositori publik Kaggle, yang menyediakan data tekanan darah sistolik (SystolicBP) dan diastolik (DiastolicBP) dari total 1.878 sampel individu. Dataset ini difokuskan pada dua fitur utama yang paling relevan dalam konteks hipertensi, yaitu nilai tekanan darah sistolik dan diastolik, tanpa melibatkan fitur tambahan lainnya. Sebelum dilakukan klasterisasi, data dibersihkan dari duplikasi untuk memastikan tidak terjadi bias akibat pengulangan entri yang sama. Selanjutnya, dilakukan standarisasi fitur menggunakan metode StandardScaler untuk memastikan bahwa kedua variabel memiliki skala yang setara, mengingat K-Means sensitif terhadap skala data. Standarisasi ini bertujuan agar proses klasterisasi berjalan optimal dan tidak terpengaruh oleh perbedaan satuan atau rentang nilai antar fitur.

2.2. K-Means *Clustering*

Penelitian ini menerapkan algoritma K-Means Clustering untuk mengelompokkan individu ke dalam klaster berdasarkan kemiripan nilai tekanan darah. K-Means merupakan metode pembelajaran tanpa pengawasan (unsupervised learning) yang secara iteratif menyempurnakan posisi centroid guna meminimalkan variansi intra-klaster. Algoritma ini beroperasi secara iteratif untuk meminimalkan variansi dalam klaster, yang dirumuskan dalam bentuk fungsi objektif sebagai berikut:

$$J = \sum_{i=1}^{K} \sum_{x \in C_i} ||\, x - \mu_i ||^2 \qquad (1)$$

Pada persamaan (1), di mana *J* merupakan total variansi dalam klaster, *K* adalah jumlah klaster, *x* adalah titik data, dan *μi* adalah centroid dari klaster ke-*i*. Tujuan dari algoritma ini adalah untuk menemukan posisi centroid *μi* yang meminimalkan jumlah kuadrat jarak antara semua titik data dalam klaster dengan centroid-nya masing-masing. Tahapan algoritma diawali dengan inisialisasi centroid secara acak. Selanjutnya dilakukan proses *assignment*, yaitu mengalokasikan setiap data ke klaster terdekat berdasarkan jarak Euclidean, dapat dilihat pada persamaan 2:

$$c_j = \arg\min_i |\, x_j - \mu_i |^2 \qquad (2)$$

Setelah semua data dialokasikan ke klaster, langkah berikutnya adalah *update*, yakni menghitung ulang centroid dengan mencari rata-rata seluruh data dalam klaster:

$$\mu_i = \frac{1}{|C_i|} \sum_{x_j \in C_i} x_j \qquad (3)$$

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Dengan menggunakan persamaan (3) proses *assignment* dan update ini dilakukan secara berulang hingga tidak terjadi lagi perubahan signifikan pada posisi centroid atau hingga fungsi objektif *J* konvergen. Berdasarkan hasil metode Elbow, titik tekuk paling optimal diperoleh pada nilai $K = 4$, yang menunjukkan bahwa struktur alami data tekanan darah paling representatif jika dikelompokkan ke dalam empat klaster. Meskipun klasifikasi klinis hipertensi umumnya dibagi menjadi lima kategori, hasil klasterisasi K-Means merepresentasikan segmentasi statistik dari distribusi data, sehingga satu klaster dapat mencakup lebih dari satu kategori klinis yang berdekatan. Untuk menentukan jumlah klaster yang optimal, digunakan metode Elbow, yang mengamati titik tekuk (elbow point) pada grafik nilai inertia terhadap jumlah klaster. Dari hasil tersebut, dipilih jumlah klaster yang mampu merepresentasikan segmentasi alami data tekanan darah menjadi beberapa level, yang diinterpretasikan sebagai indikasi tingkat hipertensi.

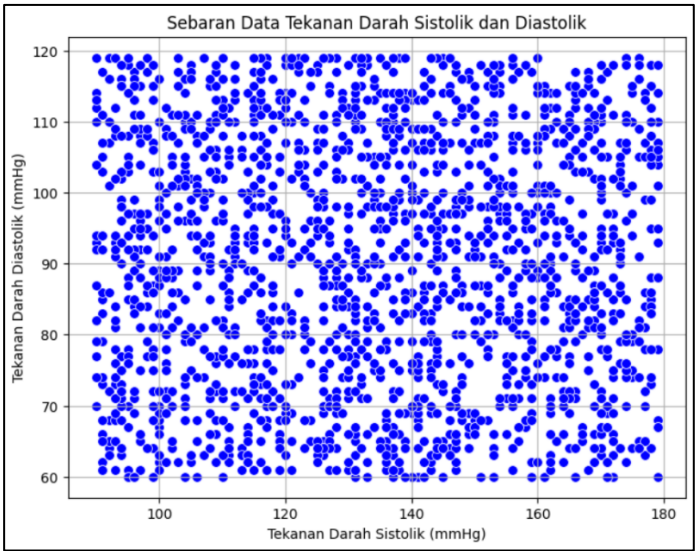
Hasil klasterisasi kemudian divisualisasikan dalam diagram sebar dua dimensi, dengan sumbu X mewakili tekanan darah sistolik dan sumbu Y mewakili tekanan darah diastolik. Setiap klaster diwakili oleh warna berbeda dan disertai centroid masing-masing, memungkinkan pengamatan visual terhadap distribusi individu berdasarkan tingkat tekanan darah mereka. Analisis deskriptif terhadap masing-masing klaster dilakukan untuk menafsirkan karakteristik tiap kelompok, misalnya: Klaster dengan nilai tekanan darah rendah cenderung menunjukkan kategori normotensi, Klaster dengan tekanan darah menengah mungkin mencerminkan prehipertensi, Klaster dengan tekanan darah tinggi berpotensi mengindikasikan hipertensi sedang atau berat. Metodologi ini tidak menggunakan label kelas eksplisit, sehingga interpretasi terhadap level hipertensi dilakukan berdasarkan pengamatan terhadap distribusi data dan centroid masing-masing klaster.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Pra-Pemrosesan Data

Tahapan awal dalam penelitian ini dilakukan melalui pemrosesan data yang diperoleh dari repositori publik Kaggle. Fokus utama ditujukan pada dua parameter vital dalam penentuan hipertensi, yaitu tekanan darah sistolik dan diastolik. Setelah dilakukan pembersihan data terhadap entri duplikat, diperoleh sebanyak 1.575 catatan unik dari total 1.878 data awal. Langkah ini penting untuk memastikan kualitas data dan menghindari bias yang mungkin timbul akibat redundansi. Data kemudian dinormalisasi menggunakan teknik StandardScaler untuk menyetarakan skala antar fitur, mengingat algoritma K-Means sangat sensitif terhadap perbedaan skala. Proses ini memastikan bahwa setiap fitur memiliki kontribusi yang seimbang dalam pembentukan klaster.

Visualisasi awal dari data sebelum klasterisasi ditampilkan pada Gambar 2, yang memperlihatkan sebaran nilai tekanan darah sistolik dan diastolik.



Gambar 2. Sebaran Data (Pra-Klaster)

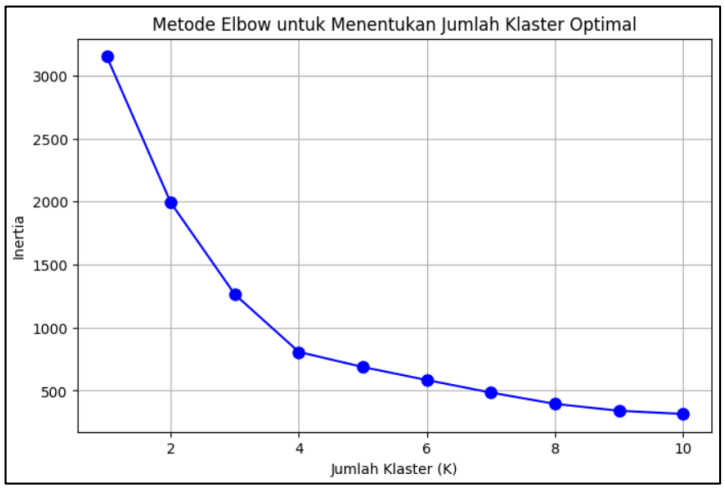
Visualisasi ini menunjukkan bahwa mayoritas data berada pada rentang sistolik antara 100–160 mmHg dan diastolik antara 70–110 mmHg. Titik-titik tersebar cukup merata, menandakan keberagaman kondisi tekanan darah pasien dalam dataset. Sebaran yang luas ini menjadi indikator bahwa metode unsupervised seperti K-Means relevan untuk mengelompokkan individu berdasarkan kemiripan profil tekanan darah. Dengan visualisasi ini, peneliti dapat

p-ISSN: 2775-4227
e-ISSN: 2775-4219

mengidentifikasi adanya kemungkinan pola-pola alami dalam data yang tidak tampak secara eksplisit. Langkah ini menjadi dasar dalam pemilihan fitur utama untuk proses klasterisasi dan pembentukan model prediksi level hipertensi berbasis unsupervised learning.

3.2. Penentuan Jumlah Klaster

Sebelum melakukan klasterisasi, jumlah klaster optimal ditentukan menggunakan metode ElbowHasil perhitungan inertia menunjukkan titik tekuk pada K = 4, yang diinterpretasikan sebagai jumlah klaster yang paling representatif terhadap struktur alami dalam data. Hal ini mengindikasikan adanya lima kelompok berbeda berdasarkan tekanan darah, yang secara klinis dapat dikaitkan dengan kategori seperti hipotensi, normotensi, prehipertensi, hipertensi tahap 1, dan hipertensi tahap 2.



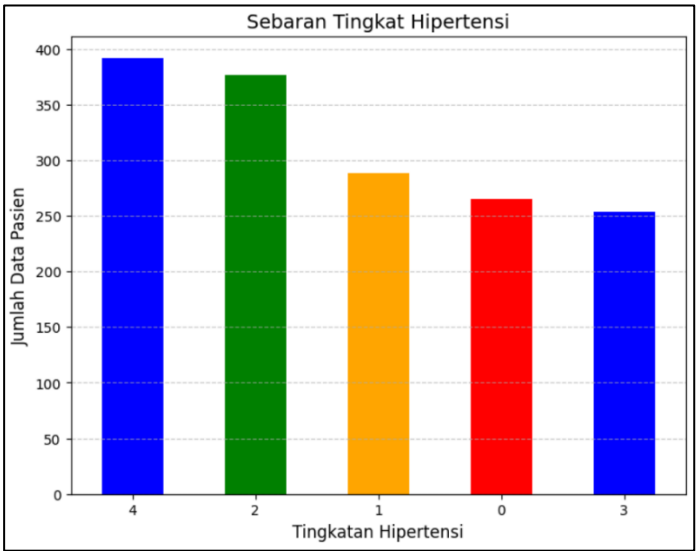
Gambar 3. Grafik Elbow

Gambar 3 menunjukkan hasil metode Elbow yang digunakan untuk menentukan jumlah klaster optimal dalam algoritma K-Means. Pada grafik tersebut, terlihat penurunan inertia yang signifikan hingga titik K = 4, setelah itu penurunan menjadi lebih landai. Hal ini mengindikasikan bahwa jumlah klaster optimal yang merepresentasikan struktur alami data adalah sebanyak empat klaster. Oleh karena itu, pemodelan akhir dalam penelitian ini menggunakan K = 4 sebagai parameter jumlah klaster pada algoritma K-Means.

3.3. Hasil Klasterisasi K-Means

Setelah nilai K ditentukan, algoritma K-Means diterapkan untuk mengelompokkan individu berdasarkan kemiripan pola tekanan darah. Hasil klasterisasi divisualisasikan dalam Gambar 4, yang menunjukkan distribusi lima level intepretasi klinis yang terbentuk di ruang dua dimensi antara tekanan darah sistolik dan diastolik. Setiap warna pada grafik menunjukkan anggota dari klaster yang berbeda, sedangkan titik pusat menunjukkan posisi centroid masing-masing klaster.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

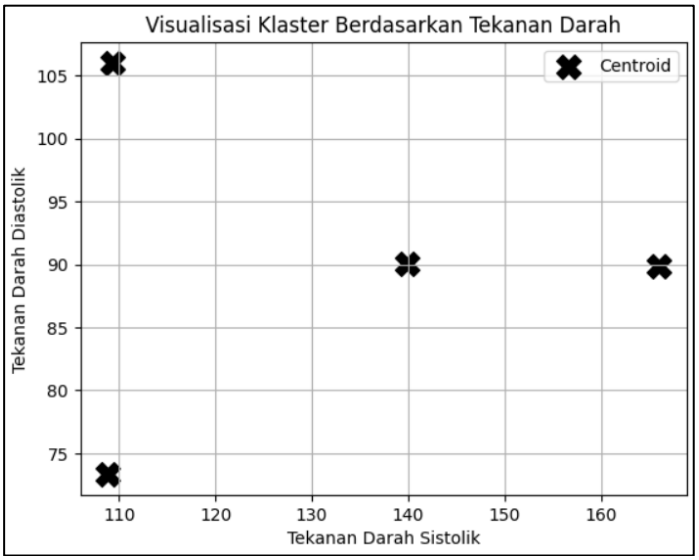


Gambar 4. Visualisasi K-Means (warna + centroid)

Analisis deskriptif terhadap masing-masing klaster menunjukkan bahwa:

- Klaster 0 memiliki tekanan darah paling rendah, kemungkinan mengindikasikan kelompok dengan hipotensi ringan.
- Klaster 1 mendekati rentang tekanan darah normal.
- Klaster 2 menunjukkan tekanan darah di batas atas normal, yang sesuai dengan kondisi prehipertensi.
- Klaster 3 dan 4 memperlihatkan tekanan darah yang meningkat signifikan, mengarah pada klasifikasi hipertensi tahap 1 dan tahap 2.

Informasi ini menunjukkan bahwa K-Means mampu mengelompokkan individu berdasarkan kecenderungan fisiologis tekanan darah tanpa memerlukan label klasifikasi, yang sangat berguna untuk mendukung sistem diagnosis awal berbasis data. Kemudian visualisasi klasterisasi dilakukan untuk memahami pola pengelompokan pasien berdasarkan tekanan darah sistolik dan diastolik.



Gambar 5. Klasterisasi K-Means Berdasarkan Tekanan Darah

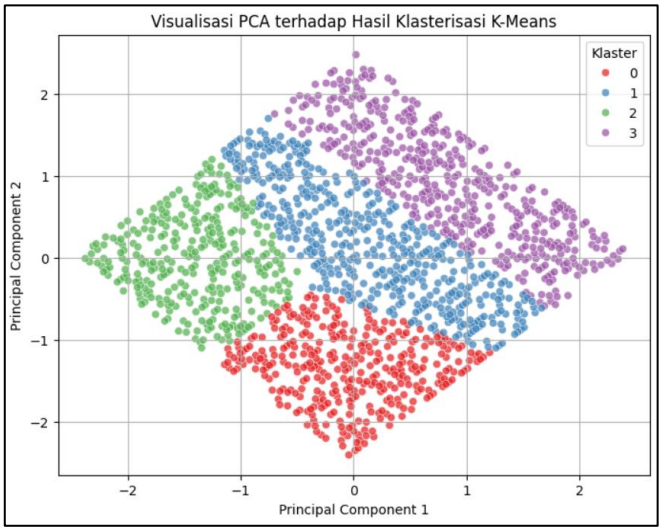
Gambar 5 menunjukkan hasil penerapan algoritma K-Means dengan jumlah klaster optimal sebanyak empat (K = 4). Setiap warna pada grafik merepresentasikan satu klaster yang mencerminkan karakteristik tekanan darah yang serupa antar individu. Titik-titik yang tersebar menggambarkan data pasien, sedangkan simbol 'X' hitam menandai posisi centroid masing-masing klaster. Pengelompokan ini memperlihatkan adanya pemisahan yang jelas antara kelompok dengan tekanan darah rendah, normal, tinggi, dan sangat tinggi. Klaster yang berada di sisi kiri bawah grafik cenderung mewakili pasien dengan tekanan darah sistolik dan diastolik yang relatif rendah,

p-ISSN: 2775-4227
e-ISSN: 2775-4219

sedangkan klaster yang terletak di sisi kanan atas menggambarkan kelompok pasien dengan tekanan darah tinggi, yang secara klinis berpotensi mengalami hipertensi derajat sedang hingga berat. Visualisasi ini mendukung interpretasi bahwa algoritma K-Means mampu mengenali pola alami dalam data tanpa adanya label awal, dan menghasilkan segmentasi yang bermakna terhadap populasi berdasarkan profil tekanan darah. Temuan ini berpotensi digunakan dalam sistem pendukung keputusan medis untuk mengklasifikasikan tingkat risiko hipertensi secara lebih awal dan berbasis data objektif.

3.4. Validasi dan Interpretasi Hasil

Untuk memperoleh pemahaman yang lebih baik terhadap hasil klasterisasi K-Means yang telah dilakukan sebelumnya, diterapkan teknik Principal Component Analysis (PCA) guna mereduksi dimensi data ke dalam dua komponen utama. PCA memungkinkan representasi visual dari data berdimensi tinggi menjadi dua dimensi yang dapat divisualisasikan, tanpa mengurangi karakteristik penting dari sebaran data. Gambar 6 menunjukkan hasil visualisasi klasterisasi menggunakan PCA berdasarkan dua komponen utama. Setiap titik pada grafik merepresentasikan satu individu dalam dataset, sementara warna berbeda menunjukkan hasil pembagian klaster oleh algoritma K-Means untuk *K=4*. Keempat klaster yang terbentuk tampak terdistribusi dengan baik dan relatif terpisah satu sama lain, yang menunjukkan bahwa model K-Means berhasil mengelompokkan data ke dalam kelompok yang memiliki karakteristik berbeda secara signifikan.



Gambar 6. Visualisasi PCA terhadap Hasil Klasterisasi

Dari visualisasi ini terlihat bahwa klaster 0 (merah), klaster 1 (hijau), klaster 2 (biru), dan klaster 3 (ungu) membentuk pola distribusi yang cukup teratur, mengindikasikan adanya pemisahan yang jelas antar klaster. Hal ini memperkuat temuan sebelumnya dari metode elbow bahwa pemilihan jumlah klaster optimal sebanyak empat sudah cukup representatif untuk menggambarkan variasi data tekanan darah yang digunakan. Lebih lanjut, distribusi antar klaster yang tidak tumpang tindih secara signifikan juga mendukung efektivitas dari pendekatan hybrid yang digunakan dalam penelitian ini, yaitu dengan menggabungkan metode unsupervised (K-Means) dengan metode supervised (XGBoost) untuk klasifikasi akhir. Dengan adanya pemisahan klaster yang baik, model klasifikasi memiliki peluang lebih besar dalam melakukan pelabelan yang akurat berdasarkan karakteristik tekanan darah dan fitur klinis lainnya.

3.5. Pembahasan

Meskipun klasifikasi klinis hipertensi secara medis umumnya dibagi menjadi lima level, yaitu hipotensi, normotensi, prehipertensi, hipertensi tahap 1, dan hipertensi tahap 2, hasil klasterisasi menggunakan algoritma K-Means dalam penelitian ini menunjukkan bahwa struktur alami data tekanan darah paling optimal direpresentasikan oleh empat klaster statistik. Keempat klaster ini tidak dimaksudkan sebagai pengganti klasifikasi klinis, melainkan sebagai segmentasi berbasis distribusi data, di mana satu klaster dapat mencakup lebih dari satu kategori klinis yang berdekatan. Penentuan jumlah klaster tersebut diperoleh berdasarkan metode Elbow yang menunjukkan titik tekuk pada K = 4, di mana penambahan jumlah klaster selanjutnya tidak menghasilkan penurunan variansi yang signifikan.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Perbedaan antara jumlah klaster statistik dan jumlah kategori klinis merupakan konsekuensi dari pendekatan unsupervised learning, di mana algoritma mengelompokkan data berdasarkan struktur distribusi numerik, bukan berdasarkan batas diagnostik yang ditetapkan secara medis. Algoritma K-Means mengelompokkan data berdasarkan kedekatan dan distribusi nilai fitur tekanan darah sistolik dan diastolik tanpa mempertimbangkan label klinis. Oleh karena itu, apabila dua level klinis memiliki pola data yang tumpang tindih atau tidak cukup berbeda secara statistik, keduanya dapat tergabung dalam satu klaster. Selain itu, ketidakseimbangan jumlah sampel pada kelompok tertentu, seperti hipotensi yang relatif jarang, dapat menyebabkan kelompok tersebut tidak teridentifikasi sebagai klaster terpisah yang signifikan.

Dengan demikian, jumlah klaster yang dihasilkan algoritma mencerminkan struktur alami data tekanan darah dan bukan merupakan pengganti klasifikasi medis. Hasil ini tetap relevan untuk tujuan eksploratif, khususnya dalam membantu segmentasi pasien berdasarkan kemiripan profil tekanan darah secara statistik. Interpretasi klinis terhadap klaster yang terbentuk tetap memerlukan keterlibatan tenaga medis untuk memvalidasi batasan dan makna klinis masing-masing klaster. Temuan penelitian ini menunjukkan bahwa algoritma K-Means efektif dalam membagi populasi pasien ke dalam kelompok tekanan darah yang berbeda secara statistik, dan berpotensi menjadi fondasi awal bagi pengembangan sistem diagnosis dini dan stratifikasi risiko hipertensi berbasis data. Pendekatan ini juga membuka peluang pengembangan aplikasi kesehatan yang bersifat prediktif dan preventif.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma K-Means Clustering dapat diimplementasikan secara efektif untuk mengelompokkan individu berdasarkan parameter tekanan darah sistolik dan diastolik melalui pendekatan pembelajaran tanpa pengawasan (unsupervised learning). Hasil klasterisasi menunjukkan bahwa struktur alami data tekanan darah paling representatif jika dibagi ke dalam empat klaster utama, yang mencerminkan segmentasi level tekanan darah dari kondisi relatif rendah hingga tinggi. Pendekatan ini memungkinkan identifikasi pola laten dalam data klinis tanpa ketergantungan pada label medis, sehingga sesuai untuk tujuan eksploratif dan sebagai dasar awal pengembangan sistem pendukung keputusan medis berbasis data. Pemanfaatan metode Elbow dalam penentuan jumlah klaster serta visualisasi menggunakan Principal Component Analysis (PCA) turut memperkuat keandalan hasil klasterisasi dengan menunjukkan pemisahan klaster yang relatif jelas. Meskipun klaster yang dihasilkan tidak secara langsung merepresentasikan batas klasifikasi klinis hipertensi, hasil ini tetap relevan untuk membantu segmentasi pasien berdasarkan kemiripan profil tekanan darah secara statistik dan objektif.

Penelitian ini memiliki keterbatasan karena menggunakan dataset publik yang belum divalidasi secara klinis. Oleh karena itu, penelitian lanjutan disarankan untuk melibatkan data dari institusi layanan kesehatan nyata, integrasi variabel klinis tambahan, serta penerapan metode evaluasi yang lebih komprehensif guna meningkatkan akurasi dan relevansi klinis model. Ke depan, pendekatan ini berpotensi dikembangkan sebagai komponen awal dalam sistem diagnosis dan stratifikasi risiko hipertensi yang lebih adaptif dan personal.

DAFTAR PUSTAKA

[1] J. Lim, J. Li, M. Zhou, X. Xiao, and Z. Xu, “Machine Learning Research Trends in Traditional Chinese Medicine: A Bibliometric Review,” *International Journal of General Medicine*, 2024. DOI: 10.2147/ijgm.s495663.

[2] A. C. Sheka, O. Adeyi, J. Thompson, B. Hameed, P. A. Crawford, and S. Ikramuddin, “Nonalcoholic Steatohepatitis: A Review,” *JAMA - Journal of the American Medical Association*. 2020. DOI: 10.1001/jama.2020.2298.

[3] Whelton PK, Carey RM, Aronow WS, Casery DE, Collins KJ, and Himmelfarb CD, “ACC/AHA/AAPA/ABC/ACPM/AGS/ APhA/ ASH/ ASPC/ NMA / PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults,” *Hypertension*, 2018. DOI: 10.1016/j.jacc.2018.03.016.

[4] Yeni Yulianti, Teten Tresnawan, Yosep Purnairawan, Susilawati, and Arneta Oktavia, “Identification Of Factors Affecting The Quality Of Life In Hypertension Patients,” *HealthCare Nursing Journal*, 2023. DOI: 10.35568/healthcare.v5i2.3307.

[5] P. Paschos and K. Paletas, “Non alcoholic fatty liver disease and metabolic syndrome,” *Hippokratia*. 2009. DOI: 10.31032/ijbpas/2021/10.1.1009.

[6] K. G. M. M. Alberti *et al.*, “Harmonizing the metabolic syndrome: A joint interim statement of the international diabetes federation task force on epidemiology and prevention; National heart, lung, and blood institute; American heart association; World heart federation; International ,” *Circulation*, vol. 120,

p-ISSN: 2775-4227
e-ISSN: 2775-4219

no. 16, pp. 1640–1645, 2009. DOI: 10.1161/CIRCULATIONAHA.109.192644.

[7] Wan K S *et al.*, "The metabolic syndrome epidemic in Malaysia 'A ticking time bomb'. Prevalence of metabolic syndrome and metabolicdysfunction-associated fatty liver disease in Malaysia 2023.," 2024.

[8] B. M. Egan, "Insulin resistance and the sympathetic nervous system," *Current Hypertension Reports*. 2003. DOI: 10.1007/s11906-003-0028-7.

[9] M. C. S. Moreira *et al.*, "Does the sympathetic nervous system contribute to the pathophysiology of metabolic syndrome?," *Frontiers in Physiology*. 2015. DOI: 10.3389/fphys.2015.00234

[10] S. Ghosh and S. Kumar, "Comparative Analysis of K-Means and Fuzzy C-Means Algorithms," *International Journal of Advanced Computer Science and Applications*, 2013. DOI: 10.14569/ijacsa.2013.040406.

[11] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Information Sciences*, 2023. DOI: 10.1016/j.ins.2022.11.139.

[12] L. Li, Q. Song, and X. Yang, "K-means clustering of overweight and obese population using quantile-transformed metabolic data," *Diabetes, Metabolic Syndrome and Obesity*, 2019. DOI: 10.2147/DMSO.S206640.

[13] M. F. Ijaz, G. Alfian, M. Syafrudin, and J. Rhee, "Hybrid Prediction Model for type 2 diabetes and hypertension using DBSCAN-based outlier detection, Synthetic Minority Over Sampling Technique (SMOTE), and random forest," *Applied Sciences (Switzerland)*, 2018. DOI: 10.3390/app8081325.

[14] N. Nuryani, T. P. Utomo, N. Wiyono, A. D. Sutomo, and S. H. Ling, "Cuffless Hypertension Detection Using Swarm Support Vector Machine Utilizing Photoplethysmogram and Electrocardiogram," *Journal of Biomedical Physics and Engineering*, 2023. DOI: 10.31661/jbpe.v0i0.2206-1504.

[15] T. Do *et al.*, "Utilizing XGBoost for the Prediction of Material Corrosion Rates from Embedded Tabular Data using Large Language Model," in *Proceedings - 2023 2023 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2023*, 2023. DOI: 10.1109/BIBM58861.2023.10385544.

[16] C. Chen, T., & Guestrin, "A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794).," *The Journal of the Association of Physicians of India*, 2016. DOI:

[17] V. P. Kovacheva, B. W. Eberhard, R. Y. Cohen, M. Maher, R. Saxena, and K. J. Gray, "Preeclampsia Prediction Using Machine Learning and Polygenic Risk Scores from Clinical and Genetic Risk Factors in Early and Late Pregnancies," *Hypertension*, 2024. DOI: 10.1161/HYPERTENSIONAHA.123.21053.

[18] F. Di Martino and F. Delmastro, "Explainable AI for clinical and remote health applications: a survey on tabular and time series data," *Artificial Intelligence Review*, 2023. DOI: 10.1007/s10462-022-10304-3.

[19] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics (Switzerland)*. 2020. DOI: 10.3390/electronics9081295.