# PENGELOMPOKAN PASIEN HIPERTENSI MENGGUNAKAN METODE *AFFINITY PROPAGATION* (STUDI KASUS: PUSKESMAS TAMBAK WEDI)

**Anissa Andiar Bhalqis, Kartika Maulida Hindrayani, Mohammad Idhom**
Sains Data, Universitas Pembangunan Nasional "Veteran" Jawa Timur
Jl. Raya Rungkut Madya, Gunung Anyar, Gununganyar, Surabaya, Jawa Timur
*kartika.maulida.ds@upnjatim.ac.id*

**ABSTRAK**

Hipertensi merupakan penyakit kronis yang berpotensi menimbulkan berbagai komplikasi serius, seperti penyakit jantung, stroke, dan gangguan ginjal, sehingga pemetaan risiko pasien menjadi penting dalam upaya pencegahan dan penanganannya. Peningkatan jumlah kasus hipertensi dari tahun ke tahun menunjukkan perlunya analisis pola kesehatan untuk memahami karakteristik pasien secara lebih komprehensif. Penelitian ini bertujuan mengelompokkan pasien hipertensi berdasarkan variabel umur, jenis kelamin, indeks massa tubuh (BMI), tekanan darah sistolik, dan tekanan darah diastolik menggunakan metode *Affinity Propagation* (AP). Metode AP merupakan teknik clustering *unsupervised* yang mampu menentukan jumlah klaster secara otomatis melalui pemilihan *exemplar* sebagai pusat klaster. Tahapan penelitian meliputi pembersihan data, penghapusan data duplikat, serta transformasi data menggunakan *Robust Scaler*. Berdasarkan pengaturan parameter terbaik, metode AP menghasilkan 12 klaster dengan nilai *Silhouette Coefficient* sebesar 0,26, yang menunjukkan kualitas pemisahan klaster cukup baik. Hasil klasterisasi ini diharapkan dapat memberikan gambaran karakteristik kelompok pasien hipertensi serta mendukung strategi deteksi dini dan penanganan medis yang lebih tepat sasaran.

***Kata kunci :*** *Pasien Hipertensi, Clustering, Affinity Propagation, Silhouette Coefficient*

## 1. PENDAHULUAN

Hipertensi atau tekanan darah tinggi adalah kondisi ketika tekanan darah di dalam pembuluh arteri meningkat melebihi batas normal [1]. Pada orang dewasa, seseorang dikategorikan hipertensi apabila tekanan sistolik mencapai ≥140 mmHg dan tekanan diastolik ≥90 mmHg, sementara tekanan darah normal berada di kisaran 120/80 mmHg [2]. Penyakit ini dikenal sebagai "*silent killer*" karena umumnya tidak menampakkan gejala yang jelas, sehingga banyak penderita baru mengetahui setelah terjadi komplikasi serius [3].

Secara global, hipertensi menjadi salah satu penyakit tidak menular dengan prevalensi tinggi. WHO pada tahun 2025, memperkirakan sekitar 1,4 miliar orang dewasa berusia 30-79 tahun di dunia menderita hipertensi, terutama di negara berpenghasilan rendah dan menengah [4]. Di Indonesia, Riskesdas 2018 mencatat peningkatan prevalensi hipertensi dari 26,4% pada tahun 2013 menjadi 34,1% pada tahun 2018 [5]. Di tingkat daerah, Kota Surabaya menempati posisi ketiga sebagai wilayah dengan jumlah kasus hipertensi terbanyak di Provinsi Jawa Timur, yaitu mencapai 102.599 kasus (45,32%) pada tahun 2017 [6].

Tingginya angka kejadian ini menimbulkan masalah kesehatan yang serius, terutama karena hipertensi berkontribusi pada berbagai penyakit serius seperti stroke, serangan jantung, gagal ginjal, dan kerusakan organ lainnya [3]. Risiko terjadinya hipertensi dipengaruhi oleh sejumlah faktor, seperti usia, genetik, tingkat aktivitas fisik, stres, serta kepatuhan dalam menjalani pengobatan [7].

Meskipun upaya deteksi dini telah banyak dilakukan, masih terdapat kesenjangan dalam penatalaksanaan hipertensi. Sebagian besar pasien baru mengakses layanan kesehatan ketika penyakit telah memasuki fase lanjut, sehingga penanganan yang diberikan menjadi kurang optimal. Selain itu, karakteristik pasien hipertensi yang beragam, baik dari aspek kondisi klinis maupun faktor risiko, menyulitkan tenaga kesehatan dalam menentukan strategi penanganan yang tepat. Permasalahan utama yang dihadapi adalah belum adanya pendekatan yang mampu mengelompokkan pasien hipertensi secara sistematis berdasarkan kesamaan karakteristik, sehingga pola penting dalam data pasien belum dimanfaatkan secara optimal. Oleh karena itu, teknik analisis data dan pembelajaran mesin banyak digunakan untuk mengidentifikasi pola tersembunyi yang mendukung pengambilan keputusan klinis [8].

Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk melakukan pengelompokan pasien hipertensi berdasarkan karakteristik dasar yang dimiliki masing-masing pasien. Melalui proses pengelompokan ini, diharapkan diperoleh gambaran mengenai pola karakteristik pasien, sehingga dapat menjadi landasan dalam merumuskan strategi penanganan serta perawatan yang lebih sesuai dengan kebutuhan. Untuk mencapai tujuan tersebut, penelitian ini menggunakan metode *clustering* agar proses pengelompokan dapat dilakukan secara objektif dan selaras dengan distribusi data pasien.

*Clustering* menjadi salah satu teknik analisis data yang sesuai untuk tujuan tersebut karena mampu mengidentifikasi pola tersembunyi tanpa memerlukan

label atau kategori awal. Metode ini bekerja dengan mengelompokkan data berdasarkan kemiripan karakteristik, sehingga tiap cluster menggambarkan kelompok pasien dengan kondisi yang serupa. Beragam metode *clustering* telah dikembangkan, seperti *K-Means*, DBSCAN, dan *hierarchical clustering*. Namun, sebagian besar metode tersebut mengharuskan penentuan jumlah *cluster* di awal, yang dapat menjadi kendala ketika struktur data tidak diketahui dengan jelas. Oleh karena itu, diperlukan pendekatan yang lebih fleksibel dalam menentukan jumlah kelompok secara otomatis.

Untuk mengatasi permasalahan tersebut, metode clustering yang digunakan dalam penelitian ini adalah *Affinity Propagation*. *Affinity Propagation* (AP) adalah algoritma *clustering* yang bekerja menggunakan mekanisme *message-passing*, yaitu proses pertukaran pesan antar titik data untuk menentukan titik mana yang paling cocok menjadi pusat *cluster* (*exemplar*). Proses pembentukan *cluster* pada AP dipengaruhi oleh beberapa parameter seperti *preferences*, *damping factor*, dan jumlah iterasi. Setelah *cluster* terbentuk, kualitas pengelompokan dievaluasi menggunakan *Silhouette Coefficient* guna memastikan bahwa struktur *cluster* yang dihasilkan benar-benar merepresentasikan pola dalam data [9].

Berdasarkan latar belakang tersebut, tingginya kasus hipertensi serta beragamnya karakteristik pasien menunjukkan perlunya pendekatan yang lebih tepat dalam memahami pola kondisi pasien. Dengan menerapkan metode *Affinity Propagation*, penelitian ini diharapkan mampu menghasilkan pengelompokan pasien hipertensi yang lebih akurat serta mendukung penyusunan strategi penanganan yang lebih efektif.

## 2. TINJAUAN PUSTAKA

### 2.1. Penelitian Terdahulu

Penelitian oleh Tika Purnama Putri dan Febriani berjudul *"Clustering Data Cuti Sakit Menggunakan Algoritma Affinity Propagation"* [9] menerapkan *Affinity Propagation* untuk mengelompokkan data cuti sakit karyawan. Hasil penelitian menunjukkan bahwa pengaturan parameter seperti nilai preferensi median, *damping factor* 0,9, dan 100 iterasi berpengaruh terhadap kualitas *cluster* yang terbentuk. Penelitian ini berhasil mengungkap pola kesehatan karyawan berdasarkan departemen, usia, dan jenis penyakit, sehingga memberikan wawasan bagi perusahaan dalam menyusun langkah preventif untuk meningkatkan kesehatan dan produktivitas pekerja.

Penelitian oleh Maulidya Rahmah, Ade Candra, dan Rahmat Widia Sembiring berjudul *"Identifikasi Predikat Hasil Pengelompokkan Data Kualitas Udara dengan Menggunakan Affinity Propagation dan Silhouette Coefficient"* [10] menerapkan *Affinity Propagation* untuk mengelompokkan data kualitas udara di Kota Pekanbaru. Data terlebih dahulu dinormalisasi dengan *Z-Score*, dan berbagai nilai *damping factor* diuji untuk melihat pengaruhnya terhadap jumlah *cluster*. Hasil penelitian menunjukkan bahwa kualitas *cluster* sangat dipengaruhi oleh pemilihan parameter, karena variasi *damping factor* menghasilkan jumlah klaster yang berbeda-beda. Temuan ini menegaskan pentingnya penentuan parameter yang tepat agar hasil *clustering* lebih optimal.

Penelitian oleh Dedy Panji Agustino dan tim berjudul *"Perbandingan Algoritma DBSCAN dan Affinity Propagation dalam Segmentasi Pelanggan Tenant Inkubator Bisnis"* [11] membahas perbandingan kinerja DBSCAN dan *Affinity Propagation* (AP) dalam pengelompokan pelanggan berdasarkan model RFM. Hasil penelitian menunjukkan bahwa AP menghasilkan *cluster* yang lebih stabil dan mudah diinterpretasikan, serta mampu membentuk empat segmen pelanggan dengan karakteristik yang berbeda. Temuan tersebut menunjukkan bahwa AP lebih efektif untuk mendukung strategi *Customer Relationship Managemen*t (CRM) perusahaan

### 2.2. *Clustering*

*Clustering* merupakan teknik analisis yang membagi data ke dalam beberapa klaster dengan mempertimbangkan kesamaan atribut antar data [12]. Analisis klaster berfokus pada megelompokkan data sehingga objek dalam klaster yang sama memiliki tingkat kemiripan yang tinggi, sementara objek dalam klaster yang berbeda memiliki karakteristik yang relatif berbeda [13]. Berdasarkan karakteristik tersebut, *clustering* digolongkan sebagai metode *unsupervised learning* karena proses pengelompokannya dilakukan tanpa label kelas. *Clustering* juga mampu mengidentifikasi pola atau struktur tersembunyi dalam suatu data sehingga dapat menghasilkan informasi yang penting. Selain itu, metode ini juga dapat diterapkan secara luas pada berbagai bidang, seperti klasifikasi, pengolahan citra, dan sistem pengenalan [14].

### 2.3. *Affinity Propagation* (AP)

*Affinity Propagation* (AP) merupakan algoritma *clustering* yang mudah digunakan, memiliki kecepatan tinggi, serta tingkat kesalahan yang rendah [15]. Selain itu, algoritma ini tidak memerlukan penentuan pusat *cluster* di awal, sehingga mengurangi terjadinya optimasi lokal yang disebabkan oleh pemilihan parameter yang kurang tepat. Untuk menentukan pusat *cluster* yang paling sesuai, AP menggunakan mekanisme *message-passing*, yaitu proses pertukaran pesan antar titik data [16]. Eksemplar pada AP selalu berupa titik data aktual, berbeda dengan *K-Means* yang menggunakan *centroid* hasil perhitungan. Setiap titik data dianggap sebagai calon pusat dan pesan yang dikirim selama iterasi menunjukkan tingkat kesesuaian (*affinity*) antar titik hingga diperoleh eksemplar yang paling representatif. Dengan cara tersebut, AP dapat membentuk *cluster* secara otomatis tanpa memerlukan penentuan jumlah *cluster* di awal [17].

## 3. METODE PENELITIAN

Dalam penelitian ini, proses analisis dilakukan melalui beberapa tahapan yang disusun secara sistematis. Tahapan ini mencakup pengumpulan data, persiapan data, penerapan model, dan evaluasi hasil model. Pada tahap awal, analisis statistik deskriptif digunakan untuk menggambarkan karakteristik umum data melalui ukuran atau metrik tertentu [18]. Setiap langkah dirancang untuk memastikan bahwa proses pengolahan dan pengelompokkan data berlangsung secara terstruktur sehingga hasil akhir dapat menggambarkan karakteristik data secara akurat.

### 3.1. Pengumpulan Data

Tahapan pertama dalam penelitian ini adalah proses pengumpulan data. Data yang digunakan merupakan data sekunder yang bersumber dari rekam medis pasien di Puskesmas Tambak Wedi Surabaya. Dataset mencakup informasi pasien hipertensi tahun 2024 serta seluruh pasien yang tercatat pada periode Januari hingga Maret 2024. Variabel yang dianalisis dalam penelitian ini meliputi ($X_1$) umur, ($X_2$) jenis kelamin, ($X_3$) *Body Mass Index* (BMI), ($X_4$) tensi atas (sistolik), dan ($X_5$) tensi bawah (diastolik).

Tabel 1. Data Pasien Hipertensi

| X1 | X2 | X3 | X4 | X5 |
|---|---|---|---|---|
| 58 | L | 26.562 | 129 | 93 |
| 51 | P | 25.390 | 150 | 80 |
| 64 | L | 25.390 | 126 | 80 |
| 33 | P | 28.889 | 140 | 90 |
| 58 | L | 23.140 | 116 | 78 |

### 3.2. Persiapan Data

Tahap selanjutnya dalam penelitian ini adalah persiapan data. Tahap ini bertujuan untuk memastikan bahwa data yang digunakan dalam proses analisis berada dalam kondisi bersih, terstruktur, dan siap diolah. Tahapan persiapan data berikut yang akan dilakukan:

a. Pembersihan Data
   Pada tahap pembersihan data, dilakukan serangkaian proses untuk memastikan bahwa data yang digunakan berada dalam kondisi optimal sebelum dianalisis lebih lanjut. Langkah pertama adalah memeriksa *missing value* pada setiap variabel, kemudian menangani nilai hilang tersebut dengan metode yang sesuai agar tidak menurunkan kualitas analisis. Selanjutnya, dilakukan pengecekan terhadap data duplikat yang berpotensi menyebabkan bias, dan jika ditemukan maka baris duplikat akan dihapus. Proses berikutnya adalah mengidentifikasi dan menangani outlier menggunakan teknik yang relevan agar nilai ekstrem tidak memengaruhi hasil pengelompokan. Selain itu, apabila terdapat variabel kategorikal, dilakukan proses encoding untuk mengubah kategori menjadi bentuk numerik agar dapat diproses oleh algoritma *clustering*. Proses ini dilakukan untuk memastikan dataset berada dalam kondisi bersih dan layak digunakan pada tahap selanjutnya.
b. Transformasi Data
   Transformasi data yang akan digunakan pada penelitian ini adalah *robust scaler*. *Robust Scaler* merupakan salah satu teknik *feature scaling* yang digunakan dalam analisis data untuk mengubah skala variabel agar lebih tahan terhadap keberadaan *outlier* atau nilai ekstrem. Keberadaan *outlier* dapat memengaruhi perhitungan statistik seperti rata-rata dan simpangan baku, sehingga berpotensi memberikan dampak besar terhadap hasil analisis [19]. Berikut rumus *robust scaler* ditunjukkan pada persamaan (1).

$$z = \frac{x - Q_2}{Q_3 - Q_1} \quad (1)$$

Keterangan,
$z$ : Hasil Transformasi
$x$ : Nilai awal *feature*
$Q_2$ : Median dari *feature*
$Q_3$ : Kuartil ketiga dari *feature*
$Q_1$ : Kuartil pertama dari *feature*

### 3.3. *Affinity Propagation* (AP)

Pada tahap awal dalam algoritma *Affinity Propagation*, perhitungan similarity dilakukan untuk menentukan kemiripan antar data. *Similarity* dihitung menggunakan negatif jarak Euclidean kuadrat**,** yang didefinisikan oleh rumus persamaan (2).

$$S_{(i,j)} = -\sum_{k=1}^{n} \left(x_{ik} - x_{jk}\right)^2 \quad (2)$$

Keterangan,
$S_{(i,j)}$ : Nilai kesamaan (*similarity*) antara titik data i dan j
$x_{ik}$ : Nilai pada dimensi ke-k untuk titik data ke-i
$x_{jk}$ : Nilai pada dimensi ke-k untuk titik data ke-j
$n$ : Jumlah total dimensi atau variabel dalam data

Setelah *similarity* dihitung, nilai *responsibility* diperbarui untuk menilai seberapa cocok suatu titik data menjadi eksemplar dibanding kandidat lainnya. Pada tahap ini, setiap data mengevaluasi calon eksemplar berdasarkan nilai *similarity* yang telah dihitung seperti yang ditunjukkan persamaan (3).

$$r(i,k) \leftarrow s(i,k) - max_{k' \neq k}\{a(i,k') + s(i,k')\} \quad (3)$$

Keterangan,
$r(i,k)$ : Nilai *responsibility* untuk $X_k$ sebagai pusat *cluster* bagi $X_i$
$s(i,k)$ : Nilai *similarity* antara $X_k$ dan $X_i$
$a(i,k')$: *Availability* dari kandidat $X_{k'}$ terhadap $X_i$
$s(i,k')$: *Similarity* antara $X_i$ dan kandidat eksemplar lain $X_{k'}$

Selanjutnya, *responsibility* untuk kandidat eksemplar terhadap dirinya sendiri dihitung seperti pada persamaan (4).

$$r(k,k) \leftarrow s(k,k) - max_{k' \neq k}\{s(k,k')\} \quad (4)$$

Keterangan,

$r(k,k)$ : *Responsibility* untuk $X_k$ sebagai eksemplar bagi dirinya sendiri

$s(k,k)$ : Nilai kesamaan atau *self-similarity* titik data $X_k$

$s(k,k')$ : *Similarity* antara $X_k$ dan kandidat lain $X_{k'}$

Selanjutnya, *availability* diperbarui untuk menilai sejauh mana suatu data dapat memilih eksemplar, dipengaruhi oleh *responsibility* sebelumnya. Dapat dihitung menggunakan rumus persamaan (5).

$$a(i,k) \leftarrow \min\{0, r(k,k) + \sum_{i' \neq \{i,k\}} \max\{0, r(i',k)\}\} \quad (5)$$

Keterangan,

$a(i,k)$ : *Availability* $X_k$ sebagai pusat klaster bagi $X_i$

$r(k,k)$ : *Responsibility* $X_k$ terhadap dirinya sediri sebagai pusat klaster

$r(i',k)$ : Tingkat keseuaian kandidat $X_k$ sebagai pusat klaster bagi $X_{i'}$

Selanjutnya, *availability* untuk kandidat eksemplar terhadap dirinya sendiri dihitung seperti pada persamaan (6).

$$a(k,k) = \sum_{i' \neq \{i,k\}} \max\{0, r(i',k)\}\} \quad (6)$$

Keterangan,

$a(k,k)$ : *Availability* untuk $X_k$ sebagai eksemplar bagi dirinya sendiri

$r(i',k)$ : *Responsibility* dari titik lain $X_{i'}$ terhadap kandidat eksemplar $X_k$

Tahap selanjutnya adalah *criterion* yang digunakan untuk mengevaluasi stabilitas eksemplar. Jika perubahan *responsibility* dan *availability* stabil dalam beberapa iterasi, algoritma dianggap konvergen. Dapat dihitung seperti pada persamaan (7).

$$c(i,k) = a(i,k) + r(i,k) \quad (7)$$

Keterangan,

$c(i,k)$ : Gabungan *availability* dan *responsibility* antara $X_i$ dan calon pusat klaster $X_k$

$a(i,k)$ : *Availability* yang menggambarkan kelayakan $X_k$ sebagai pusat klaster bagi $X_i$

$r(i,k)$ : *Responsibility* yang menunjukkan kesesuaian $X_k$ untuk berperan sebagai pusat klaster bagi $X_i$

Jika telah mencapai konvergensi, maka proses ini telah selesai dan setiap data akan dikelompokkan ke dalam *cluster* berdasarkan eksemplar yang telah ditentukan. Eksemplar tiap data dihitung seperti pada persamaan (8).

$$k \leftarrow \arg\max\{r(i,k) + a(i,k)\} \quad (8)$$

Keterangan,

arg max : Operator untuk menentukan nilai k yang memberikan hasil maksimum.

$r(i,k)$ : *Responsibility* yang menunjukkan kesesuaian $X_k$ untuk berperan sebagai pusat klaster bagi $X_i$

$a(i,k)$ : *Availability yang* menggambarkan kelayakan $X_k$ sebagai pusat klaster bagi $X_i$

### 3.4. Evaluasi Model

Tahap akhir penelitian ini adalah evaluasi model. Setelah pengelompokan dengan metode *Affinity Propagation* diperoleh, kualitas klaster dinilai menggunakan *silhouette coefficient*. *Silhouette coefficient* digunakan untuk mengevaluasi kualitas hasil pengelompokan (*clustering*) dengan rentang nilai antara -1 hingga 1**.** Semakin mendekati nilai 1, semakin baik tingkat kesesuaian data terhadap klaster yang terbentuk serta semakin jelas pemisahannya dari klaster lain [20]. Metode ini mengevaluasi kualitas *clustering* dengan mempertimbangkan tingkat kohesi dalam klaster dan separasi antar klaster [21]. Selain itu, *silhouette width* tiap klaster ditentukan berdasarkan nilai rata-rata *silhouette coefficient* dari seluruh data yang termasuk dalam klaster tersebut [22].

Tahap pertama dalam menghitung *silhouette coefficient* adalah menghitung jarak rata-rata dari data ke-i ke seluruh data lain dalam klaster yang sama menggunakan persamaan (9).

$$a(i) = \frac{1}{|A|} \sum_{j \in A, j \neq i} d(i,j) \quad (9)$$

Keterangan,

$a(i)$ : Rata-rata jarak data ke-i terhadap anggota lain dalam klaster A

$|A|$ : Total jumlah anggota dalam klaster A

$d(i,j)$ : Jarak antara data ke-i dan data ke-j

Selanjutnya adalah menghitung rata-rata jarak antara data ke-i dengan seluruh data pada klaster lain, misalnya C menggunakan persamaan (10).

$$d(i,C) = \frac{1}{|C|} \sum_{j \in C} d(i,j) \quad (10)$$

Keterangan,

$d(i,C)$: Rata-rata jarak antara data ke-i dengan seluruh data di klaster C

$|C|$ : Jumlah anggota dalam klaster C

$d(i,j)$ : Jarak antar data ke-i dan data ke-j

Kemudian menentukan nilai b(i) yaitu jarak rata-rata terkecil dari semua nilai d(i,C) pada klaster C yang berbeda dengan klaster A dihitung dengan persamaan (11).

$$b(i) = min_{C \neq A} d(i,C) \quad (11)$$

Keterangan,

$b(i)$ : Jarak rata-rata terkecil ke klaster C

$d(i,C)$ : Rata-rata jarak dari data ke-i ke klaster C

Tahap terakhir adalah menghitung nilai *silhouette coefficient* seperti pada persamaan (12) berikut:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (12)$$

Keterangan,
$s(i)$ : *Silhouette Coefficient* data ke-i
$b(i)$ : Rata-rata jarak ke anggota klaster sendiri
$a(i)$ : Rata-rata jarak ke klaster terdekat lain

## 4. HASIL DAN PEMBAHASAN

### 4.1. Pembersihan Data

Tahap ini akan dilakukan pembersihan data yang dilakuan dengan beberapa langkah, meliputi pengecekan dan penanganan *missing value*, penghapusan dan pengecekan data duplikat, serta *outlier*.

Tabel 2. Pengecekan *Missing Value*

| Variabel | Nilai Hilang |
|---|---|
| $X_1$ | 0 |
| $X_2$ | 0 |
| $X_3$ | 56 |
| $X_4$ | 58 |
| $X_5$ | 38 |

Tabel 1 menunjukkan terdapat *missing value* pada beberapa variabel, khususnya $X_3$, $X_4$, dan $X_5$. *Missing value* tersebut kemudian diatasi dengan metode imputasi menggunakan median karena lebih aman saat terdapat data yang sangat besar atau sangat kecil, sehingga imputasi menjadi lebih stabil.

Tabel 3. Imputasi *Missing Value*

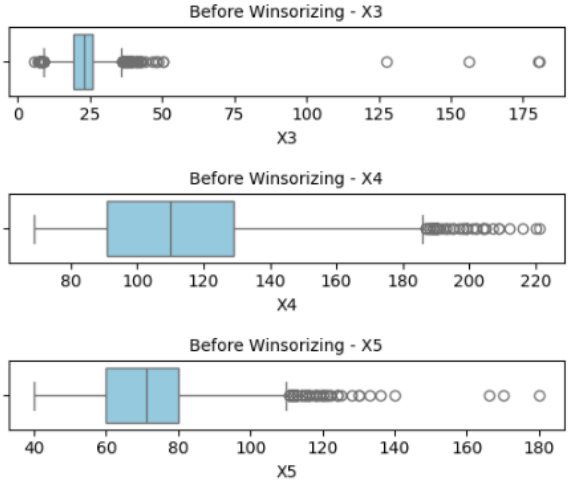
| Variabel | Nilai Hilang |
|---|---|
| $X_1$ | 0 |
| $X_2$ | 0 |
| $X_3$ | 0 |
| $X_4$ | 0 |
| $X_5$ | 0 |

Tabel 2 menunjukkan hasil penanganan *missing value*, di mana seluruh variabel sudah tidak terdapat nilai hilang. Selanjutnya dilakukan pengecekan dan penghapusan data duplikat untuk memastikan setiap entri dalam dataset bersifat unik.

```
Jumlah baris duplikat: 718
```

Gambar 1. Data Duplikat

Gambar 2 menunjukkan hasil data duplikat sebanyak 718 baris. Semua duplikat tersebut akan dihapus sehingga dataset menjadi lebih bersih dan setiap entri bersifat unik. Setelah itu identifikasi *outlier* untuk memastikan distribusi data lebih stabil.




Gambar 2. Identifikasi *Outlier*

Gambar 3 menunjukkan adanya nilai ekstrem (outlier) pada kolom numerik tersebut. Namun, pada tahap ini outlier tidak dihapus atau diubah, melainkan hanya diidentifikasi untuk memahami pola penyebarannya. Hal ini karena nilai ekstrem tersebut berpotensi merepresentasikan kondisi klinis pasien yang memang berada pada rentang tekanan darah sangat tinggi atau rendah, sehingga tetap dianggap sebagai bagian dari variasi alami data.

### 4.2. Transformasi Data

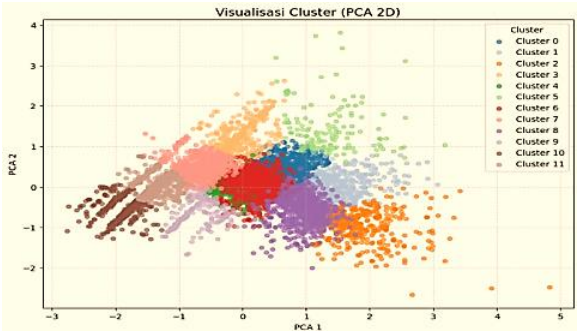
Pada penelitian ini digunakan metode *Robust Scaler*, yaitu teknik transformasi data yang memanfaatkan median serta rentang antar-kuartil (IQR) sebagai dasar perhitungannya. Tabel 4 merupakan contoh data sebelum dan sesudah dilakukan transformasi data.

Tabel 4. Hasil Transformasi Data

| **X3** | |
|---|---|
| **Data Asli** | ***Z-Score*** |
| 26.562 | 0.540 |
| 25.390 | 0.367 |
| 25.390 | 0.882 |
| 28.888 | 0.036 |
| 23.140 | -1.144 |

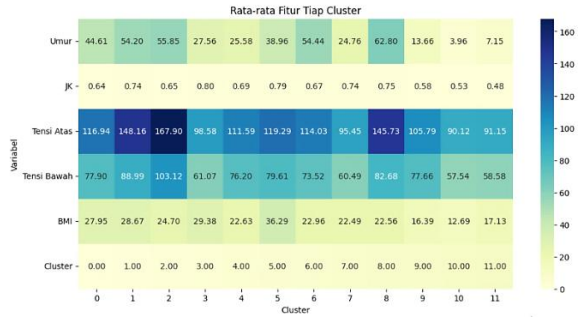
### 4.3. Pemodelan Affinity Propagation dan Evaluasi

Pada tahap pemodelan, metode *Affinity Propagation* diterapkan untuk mengelompokkan data pasien hipertensi yang telah melalui proses transformasi. Metode ini secara otomatis menentukan pusat klaster (*exemplar*) berdasarkan nilai *similarity* antar data, yang dihitung sebagai kebalikan dari kuadrat jarak Euclidean. Nilai *preference*, yang menunjukkan seberapa besar kemungkinan suatu data menjadi pusat klaster, ditetapkan menggunakan nilai minimum dari matriks *similarity*, yaitu -77,289. Selain itu, digunakan *damping factor* sebesar 0,9 untuk menjaga kestabilan konvergensi algoritma sehingga pertukaran pesan antar data berjalan lebih stabil dan tidak berfluktuasi secara berlebihan. Visualisasi klaster ditunjukkan pada Gambar 3.




Gambar 3. Hasil Visualisasi Cluster

Gambar 3 menunjukkan hasil pengelompokan pasien hipertensi menggunakan Affinity Propagation yang direduksi ke dua dimensi dengan PCA. Sumbu X (PCA 1) mewakili komponen dengan variasi terbesar, sedangkan sumbu Y (PCA 2) menggambarkan sumber variasi terbesar kedua. Setiap titik merepresentasikan satu pasien, dan perbedaan warna menunjukkan klaster yang terbentuk. Dari visualisasi terlihat bahwa beberapa klaster tampak cukup kompak, menandakan kemiripan karakteristik pasien dalam kelompok tersebut. Sementara itu, beberapa klaster tampak saling beririsan, mengindikasikan adanya kemiripan pola antar kelompok sehingga pemisahannya tidak sepenuhnya tegas.




Gambar 4. Karakteristik Tiap Cluster

Gambar 4 menampilkan *heatmap* yang memperlihatkan rata-rata setiap fitur kesehatan pada masing-masing klaster hasil pengelompokan menggunakan metode *Affinity Propagation*. Warna yang semakin gelap menunjukkan nilai rata-rata yang lebih tinggi pada suatu variabel. Dari gambar tersebut, terlihat bahwa variabel tekanan darah sistolik (Tensi Atas) dan diastolik (Tensi Bawah) menjadi faktor pembeda yang signifikan antar klaster, misalnya klaster 1, 2, dan 3 memiliki nilai Tensi Atas yang tinggi, sementara Cluster 10 dan 11 menunjukkan nilai tensi yang lebih rendah. Variabel umur juga menunjukkan perbedaan antar klaster, di mana beberapa cluster seperti klaster 6 memiliki rata-rata umur lebih tinggi dibanding cluster lainnya. Variabel jenis kelamin (JK) relatif stabil di sebagian besar cluster, menunjukkan distribusi gender yang merata. Selain itu, rata-rata BMI menunjukkan variasi yang berbeda pada beberapa klaster, menandakan perbedaan kondisi fisik pasien.

## 5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa penerapan metode *Affinity Propagation* berhasil mengelompokkan pasien hipertensi berdasarkan karakteristik dasar seperti umur, jenis kelamin, BMI, serta tekanan darah sistolik dan diastolik. Hasil pengelompokan membentuk beberapa cluster dengan perbedaan yang jelas pada variabel tekanan darah dan umur, sementara distribusi jenis kelamin relatif merata dan BMI menunjukkan variasi antar cluster. Nilai *Silhouette Coefficient* sebesar 0,26. Hasil klasterisasi ini memberikan gambaran yang lebih jelas mengenai profil pasien hipertensi, sehingga dapat mendukung tenaga kesehatan dalam merumuskan strategi penanganan dan perawatan yang lebih tepat sesuai kebutuhan masing-masing kelompok pasien. Penelitian selanjutnya disarankan dapat ditingkatkan dengan menambahkan variabel klinis yang lebih detail atau menerapkan optimasi parameter pada *Affinity Propagation* agar jumlah klaster lebih stabil dan representatif.

## DAFTAR PUSTAKA

[1] G. R. Jannah, N. Amanah, S. N. Holilah, R. Saputri, Y. P. Lestari, and A. R. Hakim, "Empowerment of Health Cadres in Sungai Batang Ilir Village Through Education about berkembang . Data Global Status Report on Noncommunicable Disesases dari WHO masyarakat dan semua tenaga kesehatan . Upaya yang bersumber dari masyarakat diberi bekal penge," vol. 1, pp. 176–181, 2023.

[2] E. R. Dumalang, F. Lintong, and V. R. Danes, "Analisa Perbandingan Pengukuran Tekanan Darah antara Posisi Tidur dan Posisi Duduk pada Lansia," *J. Biomedik JBM*, vol. 14, no. 1, pp. 96–101, 2022, doi: https://doi.org/10.35790/jbm.v14i1.37592.

[3] R. Ariyanti, I. A. Preharsini, and B. W. Sipolio, "Edukasi Kesehatan Dalam Upaya Pencegahan dan Pengendalian Penyakit Hipertensi Pada Lansia," *To Maega J. Pengabdi. Masy.*, vol. 3, no. 2, pp. 74–82, 2020, doi: 10.35914/tomaega.v3i2.369.

[4] "Hypertension," World Health Organization. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/hypertension

[5] Badan Penelitian dan Pengembangan Kesehatan, *Laporan Nasional Riskesdas 2018*. Jakarta: Kementerian Kesehatan, 2019. [Online]. Available: https://drive.google.com/file/d/1GHS6lCsSfhuIU_ZkUuKpWvI1mWJ1ZFPr/view

[6] D. K. Provinsi and J. Timur, *PROFIL KESEHATAN PROVINSI JAWA TIMUR TAHUN 2018*. Dinas Kesehatan Provinsi Jawa Timur, 2018. [Online]. Available: https://dinkes.jatimprov.go.id/userfile/dokumen/BUKU PROFIL KESEHATAN JATIM 2018.pdf

[7] Z. Anshari, "Komplikasi Hipertensi Dalam Kaitannya Dengan Pengetahuan," vol. 2, no. 2, 2020.

[8] P. A. Riyantoko, T. M. Fahrudin, K. M. Hindrayani, and M. Idhom, "Exploratory Data Analysis and Machine Learning Algorithms to Classifying Stroke Disease," *Ijconsist Journals*, vol. 2, no. 02, pp. 77–82, 2021, doi: 10.33005/ijconsist.v2i02.49.

[9] F. Putri, T. P., & Febriani, "CLUSTERING DATA CUTI SAKIT MENGGUNAKAN ALGORITMA AFFINITY PROPAGATION (STUDI KASUS: PERUSAHAAN TELEKOMUNIKASI DI JAKARTA)," *J. Ilm. Teknol. dan Rekayasa*, vol. 27(1), pp. 69–84, 2022.

[10] M. Rahmah, A. Candra, and R. W. Sembiring, "Identifikasi Predikat Hasil Pengelompokan Data Kualitas Udara dengan Menggunakan Affinity Propagation dan Silhouette Coefficient," *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 6, no. 2, pp. 176–180, 2022, [Online]. Available: https://core.ac.uk/download/pdf/522725028.pdf

[11] D. P. Agustino, I. G. Bintang, A. Budaya, I. G. Harsemadi, I. K. Dharmendra, and I. Made, "Perbandingan Algoritma DBSCAN dan Affinity Propagation dalam Segmentasi Pelanggan Tenant Inkubator Bisnis," vol. 12, pp. 315–321, 2023.

[12] A. PS, A. Nugroho Sihananto, and D. Arman Prasetya, "Implementasi Metode K-NN dalam Klasterisasi Kasus Kesehatan Jantung," *ALINIER J. Artif. Intell. Appl.*, vol. 3, no. 2, pp. 18–21, 2022, doi: 10.36040/alinier.v3i2.5761.

[13] A. M. Sahat Renaldi. S, Dwi Arman Prasetya, "Analisis KlasterPartitioningAround Medoidsdengan Gower Distanceuntuk Rekomendasi Indekos(Studi Kasus: Indekos di Sekitar Kampus UPNVJT)," *G-Tech J. Teknol. Terap.*, vol. 8, pp. 2060–2069, 2024.

[14] I. Virgo, S. Defit, and Y. Yuhandri, "Klasterisasi Tingkat Kehadiran Dosen Menggunakan Algoritma K-Means Clustering," *J. Sistim Inf. dan Teknol.*, vol. 2, pp. 23–28, 2020, doi: 10.37034/jsisfotek.v2i1.17.

[15] M. Izzatillah and A. B. Mutiara, "Pengujian Algoritma Clustering Affinity Propagation dan Adaptive Affinity Propagation terhadap IPK dan Jarak Rumah," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 4, no. 3, p. 330, 2020, doi: 10.30998/string.v4i3.6197.

[16] L. Wang *et al.*, "An Improved Integrated Clustering Learning Strategy Based on Three-Stage Affinity Propagation Algorithm with Density Peak Optimization Theory," *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/6666619.

[17] A. Sarica, M. G. Vaccaro, A. Quattrone, and A. Quattrone, "A novel approach for cognitive clustering of parkinsonisms through affinity propagation," *Algorithms*, vol. 14, no. 2, 2021, doi: 10.3390/a14020049.

[18] T. M. Fahrudin, A. Riyantoko, K. M. Hindrayani, G. Susrama, and M. Diyasa, "Seminar Nasional Informatika Bela Negara (SANTIKA) Exploratory Data Analysis pada Kasus COVID-19 di Indonesia Menggunakan HiveQL dan Hadoop Environment," *Semin. Nas. Inform. Bela Negara*, vol. 1, pp. 115–123, 2020.

[19] W. Nugraha, R. Sabaruddin, and S. Murni, "Teknik Scaling Menggunakan Robust Scaler Untuk Mengatasi Outlier Data Pada Model Prediksi Serangan Jantung," *Techno.Com*, vol. 23, no. 2, pp. 319–327, 2024, doi: 10.62411/tc.v23i2.10463.

[20] D. A. Prasetya, A. P. Sari, M. Idhom, and A. Lisanthoni, "Optimizing Clustering Analysis to Identify High-Potential Markets for Indonesian Tuber Exports," *Indones. J. Electron. Electromed. Eng. Med. Informatics*, vol. 7, no. 1, pp. 113–122, 2025, doi: 10.35882/skzqbd57.

[21] T. M. Fahrudin, P. A. Riyantoko, K. M. Hindrayani, and M. H. P. Swari, "Cluster Analysis of Hospital Inpatient Service Efficiency Based on BOR, BTO, TOI, AvLOS Indicators using Agglomerative Hierarchical Clustering," *Telematika*, vol. 18, no. 2, p. 194, 2021, doi: 10.31315/telematika.v18i2.4786.

[22] A. Muhaimin and K. Fithriasari, "Kohonen-SOM LOF Approach for Anomaly Detection," *Proc. - 2021 IEEE 7th Inf. Technol. Int. Semin. ITIS 2021*, pp. 8–13, 2021, doi: 10.1109/ITIS53497.2021.9791596.