BAB IV
HASIL DAN PEMBAHASAN

4.1 Data Selection

Bab ini membahas hasil implementasi algoritma K-Means untuk klasterisasi penyakit stroke berdasarkan gejala dan faktor risiko. Pembahasan mencakup proses pengolahan data, implementasi algoritma, hasil klasterisasi, analisis, serta validasi hasil. Data yang digunakan dianalisis untuk mengidentifikasi pola dan hubungan antara gejala serta faktor risiko terhadap klaster yang terbentuk.

4.2 Tahap Analisa

4.2.1 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini terdiri dari 5.110 data yang diambil dari platform Kaggle. Dataset ini berisi informasi terkait gejala dan faktor risiko stroke, Hypertension, Heart Disease, Avg glucose level, Bmi, Smoking Status, dan atribut lainnya. Dataset ini dipilih karena memiliki variasi fitur yang relevan dan jumlah data yang memadai untuk proses analisis. Dengan memanfaatkan data dari Kaggle, penelitian ini bertujuan untuk mengelompokkan data pasien berdasarkan karakteristik tersebut menggunakan algoritma K-Means, sehingga dapat diidentifikasi pola atau kelompok yang signifikan dalam hubungan antara faktor risiko dan kejadian stroke. Dataset ini mencakup informasi mengenai beberapa atribut yang relevan dengan faktor risiko stroke, antara lain:

Tabel 4. 1 Metadata

1.	Id	Tingkat stres yang dialami oleh individu.
2.	Gender	Skor yang merepresentasikan tingkat depresi yang dialami oleh individu.
3.	Age	Faktor risiko utama untuk stroke; usia tua lebih rentan terhadap penyakit ini.
4.	Hypertension	Hipertensi adalah salah satu penyebab utama stroke.
5.	Heart_disease	Faktor risiko yang signifikan untuk stroke.
6.	Ever_married	Mungkin digunakan untuk analisis gaya

		hidup, tetapi tidak secara langsung terkait dengan faktor medis stroke.
7.	Work_type	Dapat memberikan wawasan tentang tingkat aktivitas fisik atau stres, meskipun relevansinya tidak langsung.
8.	Residence_type	Bisa digunakan untuk melihat hubungan antara akses ke layanan kesehatan dan risiko stroke.
9.	Avg_glucose_level	Faktor medis yang relevan karena kadar gula darah yang tinggi berhubungan dengan diabetes, yang merupakan risiko stroke.
10.	Bmi	Faktor risiko terkait obesitas; Indeks Masa Tubuh (BMI) tinggi dapat meningkatkan risiko stroke.
11.	Smoking_status	Kebiasaan merokok merupakan salah satu faktor risiko stroke.
12.	Stroke	Label target untuk validasi hasil klasterisasi atau analisis risiko.

Berikut visualisasi data set Penyakit Stroke pada rapidminer :

id	gender	age	hypertension	heart_disea...	ever_married	work_type	Residence_t...	avg_glucos...	bmi	smoking_st...	stroke
31112	Male	80	0	1	Yes	Private	Rural	105.920	32.5	never smoked	1
60182	Female	49	0	0	Yes	Private	Urban	171.230	34.4	smokes	1
1665	Female	79	1	0	Yes	Self-employed	Rural	174.120	24	never smoked	1
56669	Male	81	0	0	Yes	Private	Urban	186.210	29	formerly smo...	1
53882	Male	74	1	1	Yes	Private	Rural	70.090	27.4	never smoked	1
10434	Female	69	0	0	No	Private	Urban	94.390	22.8	never smoked	1
27419	Female	59	0	0	Yes	Private	Rural	76.150	N/A	Unknown	1
60491	Female	78	0	0	Yes	Private	Urban	58.570	24.2	Unknown	1
12109	Female	81	1	0	Yes	Private	Rural	80.430	29.7	never smoked	1
12095	Female	61	0	1	Yes	Govt_job	Rural	120.460	36.8	smokes	1
12175	Female	54	0	0	Yes	Private	Urban	104.510	27.3	smokes	1
8213	Male	78	0	1	Yes	Private	Urban	219.840	N/A	Unknown	1

Gambar 4. 1 Sample Data Mentah

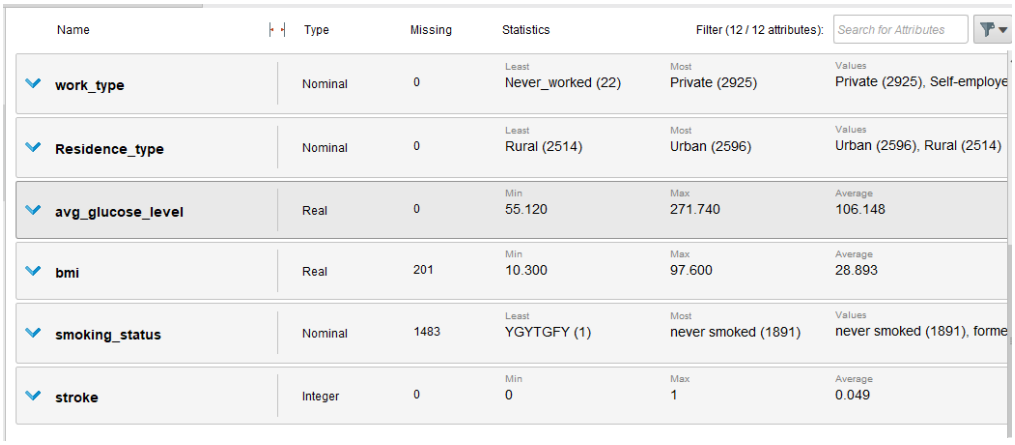
4.2.2 Proses Pra-Pemrosesan Data

Sebelum menerapkan algoritma K-Means, beberapa tahapan pra-pemrosesan dilakukan untuk memastikan data siap digunakan. Langkah-langkah ini termasuk:

a. Pembersihan Data *(Data Cleaning)*

- Missing Values

  Missing values merujuk pada data yang hilang atau tidak tersedia dalam dataset, yang dapat disebabkan oleh kesalahan pengumpulan data atau ketidaktahuan responden. Keberadaan missing values dapat memengaruhi akurasi analisis, sehingga perlu ditangani dengan metode seperti menghapus baris yang hilang atau mengisi nilai kosong dengan estimasi seperti rata-rata atau median untuk memastikan kualitas data dan hasil yang valid. Berikut Visualisasi nya :



Gambar 4. 2 Missing Values

Setelah melalui proses pembersihan data (data cleaning), ditemukan adanya missing values pada beberapa atribut sebanyak 201 pada data Bmi dan 1483 pada data Smoking Status. Dataset awal yang digunakan dalam penelitian ini terdiri dari 5110 data. Setelah melalui proses missing values data menjadi 3426 data.

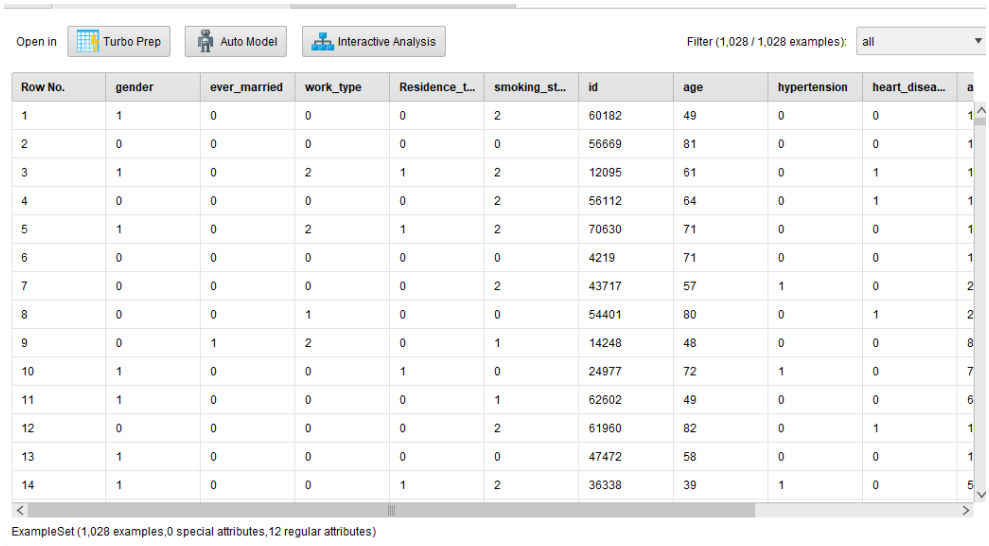
- Menghapus Duplikasi

  Hasil pengecekan menunjukkan bahwa tidak ada data yang terduplikasi dalam dataset ini. Setiap entri dalam dataset bersifat unik, sehingga memastikan keandalan dan keakuratan informasi yang tersedia. Dataset ini sudah siap untuk diproses lebih lanjut tanpa perlu penghapusan atau penyesuaian akibat duplikasi data.

b. Transformasi

- Encoding Fitur Kategorikal

  Algoritma seperti K-Means, yang menggunakan jarak (misalnya, Euclidean distance), tidak dapat langsung menangani data kategorikal. Oleh karena itu, encoding diperlukan untuk mengonversi data kategorikal menjadi numerik agar algoritma dapat menghitung jarak dengan benar. Berikut visualisasi kategorikal ke numerikal :

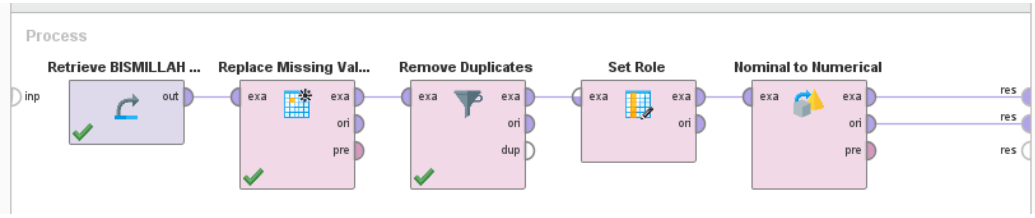


Row No.	gender	ever_married	work_type	Residence_t...	smoking_st...	id	age	hypertension	heart_disea...	a
1	1	0	0	0	2	60182	49	0	0	1
2	0	0	0	0	0	56669	81	0	0	1
3	1	0	2	1	2	12095	61	0	1	1
4	0	0	0	0	2	56112	64	0	1	1
5	1	0	2	1	2	70630	71	0	0	1
6	0	0	0	0	0	4219	71	0	0	1
7	0	0	0	0	2	43717	57	1	0	2
8	0	0	1	0	0	54401	80	0	1	2
9	0	1	2	0	1	14248	48	0	0	8
10	1	0	0	1	0	24977	72	1	0	7
11	1	0	0	0	1	62602	49	0	0	6
12	0	0	0	0	2	61960	82	0	1	1
13	1	0	0	0	0	47472	58	0	0	1
14	1	0	0	1	2	36338	39	1	0	5

ExampleSet (1,028 examples,0 special attributes,12 regular attributes)

Gambar 4. 3 Encoding Fitur Kategorikal

Berikut langkah-langkah pra-pemrosesan pada rapidminer:



Gambar 4. 4 Proses Rapidminer

c. Seleksi Fitur

Tabel 4. 2 Seleksi Fitur

Age	hypertension	heart_disease	avg_glucose_level	bmi	smoking_status
67	0	1	228.69	36.6	formerly smoked
61	0	0	202.21	N/A	never smoked
80	0	1	105.92	32.5	never smoked
49	0	0	171.23	34.4	smokes
79	1	0	174.12	24	never smoked
81	0	0	186.21	29	formerly smoked
74	1	1	70.09	27.4	never smoked
69	0	0	94.39	22.8	never smoked
59	0	0	76.15	N/A	Unknown
78	0	0	58.57	24.2	Unknown
81	1	0	80.43	29.7	never smoked
61	0	1	120.46	36.8	smokes
54	0	0	104.51	27.3	smokes
78	0	1	219.84	N/A	Unknown

Dari data mentah yang tersedia, dilakukan seleksi variabel untuk memastikan relevansi terhadap analisis yang akan dilakukan. Variabel yang dipilih meliputi age (usia), hypertension (riwayat hipertensi), avg_glucose (rata-rata kadar glukosa darah), heart_disease (riwayat penyakit jantung), smoking_status (status merokok), dan BMI (indeks massa tubuh). Pemilihan ini bertujuan untuk mengidentifikasi faktor risiko kesehatan yang dapat berkontribusi terhadap kondisi penyakit tertentu, sehingga analisis lebih lanjut dapat memberikan wawasan yang lebih akurat dan bermakna.

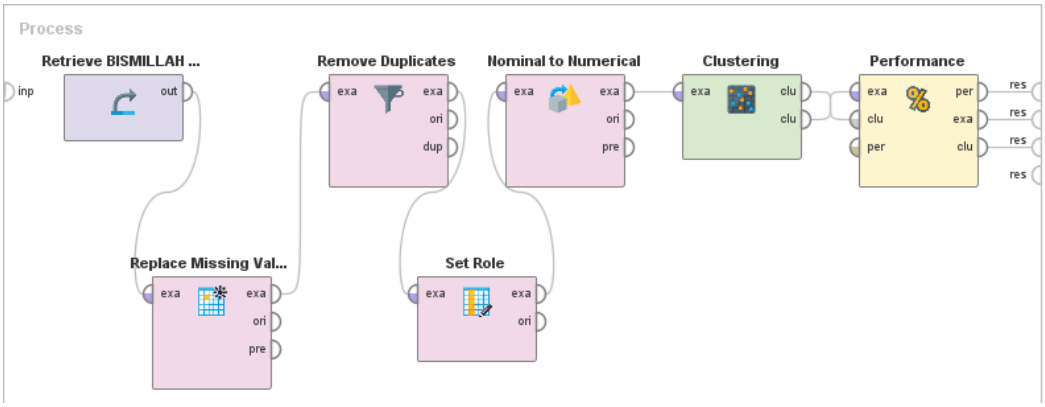
4.3 Implementasi Algoritma K-Means

4.3.1 Tahap Pemodelan Cluster

Pada tahap pemodelan klaster ini Teknik Clustering merupakan suatu pendekatan analisis data yang bertujuan untuk secara otomatis mengelompokkan data berdasarkan kesamaan karakteristik tanpa memerlukan label kelas yang telah ditentukan sebelumnya. Sebagai bagian dari metode

unsupervised learning, teknik ini memungkinkan klasifikasi data yang tidak memiliki label kelas yang jelas atau yang belum terdefinisi. Clustering sangat berguna dalam analisis data yang kompleks, di mana informasi tentang struktur kelas sulit diketahui atau tidak tersedia. Proses ini berfokus pada identifikasi pola-pola tersembunyi dalam data, dengan tujuan membentuk kelompok-kelompok yang homogen berdasarkan tingkat kesamaan anggotanya.

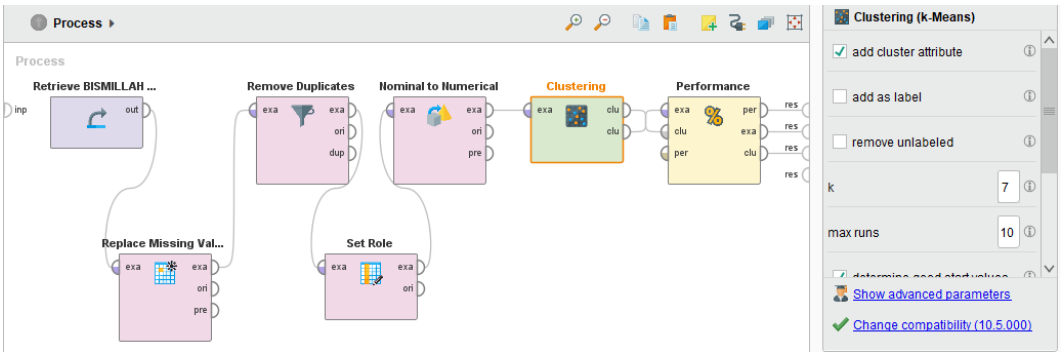
Dalam penelitian ini, metode klasterisasi yang digunakan adalah K-Means. Algoritma K-Means bekerja dengan menentukan jumlah klaster yang diinginkan (k) dan iterasi maksimum. Selain itu, algoritma K-Means bekerja dengan menunggu jumlah pengulangan eksekusi perintah, yang dikenal dengan istilah iterasi, masksimum dan sebuah nilai konstanta yang digunakan untuk penentuan kapan iterasi dihentikan. Berikut adalah implementasi algoritma K-Means dalam penelitian ini :



Gambar 4. 5 Tahap Pemodelan Cluster

Pada parameter algoritma clustering, saya melakukan simulasi sebanyak 7 kali pembentukan kelas untuk mengetahui model terbaik. Adapun max run pada parameter saya sesuaikan dengan default 10 kali untuk menentukan konsistensi dari model. Pada measures type, saya menggunakan mixed measures karena

type data pada data set yang heterogen. Berikut pengaturan pada parameter algoritma k-means perhitungan data set penyakit stroke :



Gambar 4. 6 Pengaturan Parameter

Adapun hasil dari pembentukan Cluster dari nilai k = 2 sampai k = 7 sebagai berikut:

Tabel 4. 3 Pembentukan Cluster

k-2	Hasil dari model klastering pada RapidMiner menghasilkan dua klaster utama. Cluster 0, dengan jumlah 2006 item, dan Klaster 1, dengan jumlah 392 item, memberikan gambaran mengenai pemisahan data menjadi dua kelompok yang signifikan. Klaster ini mencerminkan pola-pola atau karakteristik tertentu yang membedakan antara dua kelompok tersebut. Klaster dengan jumlah item terendah adalah Cluster 1, sedangkan Klaster 0 memiliki jumlah item tertinggi. Perbedaan jumlah item ini mengindikasikan bahwa Cluster 0 mungkin memiliki variasi atau kompleksitas yang lebih tinggi dibandingkan dengan Cluster 1
k-3	Cluster 0 dengan 710 item, dan Cluster 1 dengan 1331 item, sedangkan Cluster 2 357 item. Analisis lebih lanjut

	mengungkapkan bahwa Cluster 1 merupakan klaster tertinggi dengan jumlah item terbanyak, sedangkan Cluster 2 adalah klaster dengan jumlah item terendah
k-4	Klaster 0 menjadi klaster dengan jumlah data terbanyak, mencapai 792 item, mengindikasikan sekelompok besar data yang memiliki kesamaan karakteristik tertentu. Klaster dengan jumlah data terendah adalah Klaster 1, yang terdiri dari 349 item. Setiap klaster mewakili kelompok data yang memiliki tingkat kemiripan yang tinggi di antara anggotanya, sementara perbedaan karakteristik antar-klaster dapat diidentifikasi.
k-5	Klaster 0 membawahi sebanyak 546 entitas data, Klaster 1 terdiri dari 315 entitas data, sementara Klaster 2 mencakup 687 entitas data. Klaster 3 memiliki 173 entitas data, dan Klaster 4 terdiri dari 677 entitas data.
k-6	Cluster 0 memiliki 469 item, Cluster 1 mencakup 254 item, sementara Cluster 2 terdiri dari 496 item. Cluster 3 memiliki 348 item, Cluster 4 terdiri dari 690 item, dan Cluster 5 mencakup 141 item
k-7	Cluster 0 terdiri dari 539 item, sementara Cluster 1 memiliki 172 item. Selanjutnya, Cluster 2 mencakup 175 item, Cluster 3 terdiri dari 174 item, dan Cluster 4 memiliki 472 item. Cluster 5 menampilkan jumlah item tertinggi dengan 437, sedangkan Cluster 6 memiliki 429

4.3.1 Tahap Analisa Klaster

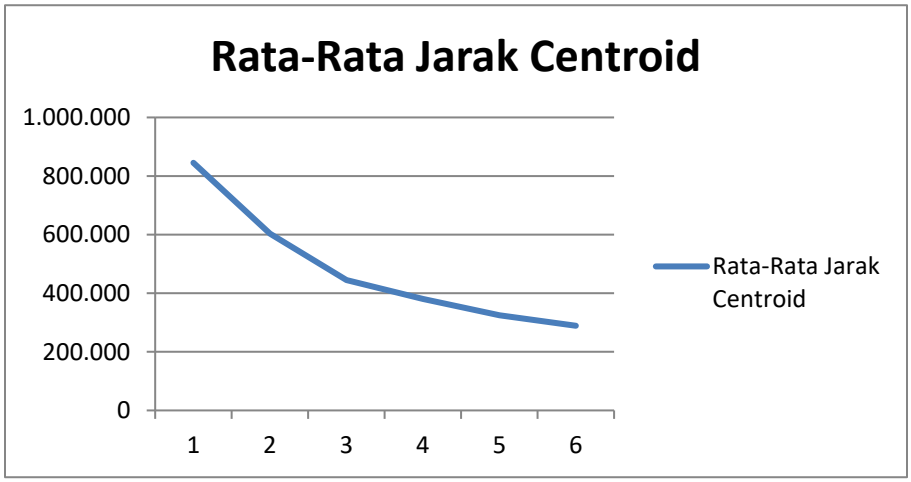
Analisis klaster merupakan salah satu teknik pada data mining yang digunakan untuk menemukan klaster-klaster dari himpunan data secara otomatis atau semi-otomatis. Berdasarkan pendekatan dan/atau konsep yang diadopsi, teknik clustering dapat dikategorikan ke dalam metoda yang berbasis partisi, hirarki,

dan densitas/kerapatan, grid, model dan konstrain. Adapun pada penelitian ini saya menggunakan simulasi model untuk menentukan klaster terbaik pada data Penyakit Stroke. Berikut hasil simulasi penentuan klaster:

Tabel 4. 4 Rata-Rata Jarak Centroid

Banyak klaster (k)	Rata-Rata Jarak Centroid
2	845.168
3	602.700
4	444.423
5	379.920
6	324.699
7	288.925

Untuk menentukan nilai klaster (k) terbaik pada data set Penyakit Stroke. Saya menggunakan metode elbow. Metode siku (Elbow Method) bertujuan untuk menemukan jumlah klaster yang optimal sehingga hasil klastering menjadi sesuai dengan struktur intrinsik dari data tanpa overfitting atau underfitting. Berikut penggunaan metode elbow (siku):



Gambar 4. 7 Rata-Rata Jarak Centroid

Berdasarkan visualisasi pada elbow method di atas, maka diketahui bahwa kelengkungan tertajam sehingga membentuk sudut siku berada pada klaster 4. Sehingga ditentukan bahwa analisis klaster terbaik berada pada klaster atau nilai k = 4.

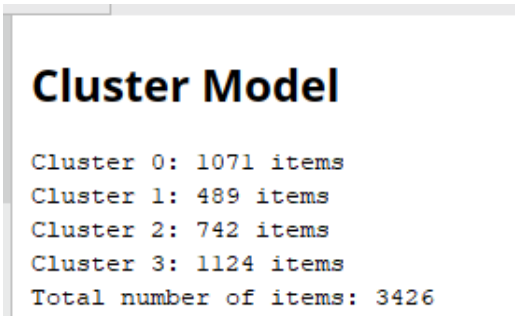
4.2.2 Kesimpulan Sementara

Berdasarkan Analisa di atas, dapat disimpulkan beberapa hasil penelitian sebagai berikut:

- Berdasarkan Analisa data set penyakit stroke memiliki nilai klaster terbaik k = 4. Adapun nilai Avg. within centroid distance adalah 444.423 nilai Avg. within centroid distance: 444.423, dapat diartikan bahwa rata-rata jarak antara titik data dalam klaster dengan pusat klaster adalah sekitar 444.423 Semakin kecil nilai ini, semakin padat dan homogen klaster tersebut.

Gambar 4. 8 Avg. within centroid distance

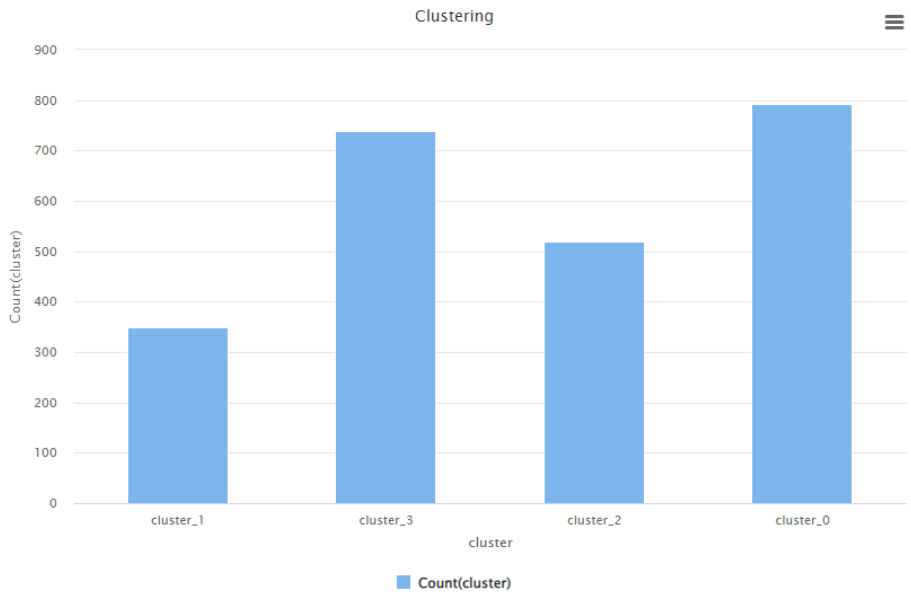
- Adapun rincian masing-masing klaster dibagi menjadi Klaster 0 dengan jumlah 1071 item, Klaster 1 dengan jumlah 489 item, Klaster 2 dengan jumlah 742 item dan Klaster 3 dengan jumlah 1124 item. Berikut gambar cluster model yang terbentuk:



Gambar 4. 9 Cluster Model

4.2.3 Visualisasi Hasil Klasterisasi

Berdasarkan hasil analisa data menggunakan metode k-means maka diperoleh bahwa jumlah kelas terbaik adalah Ketika membagi data menjadi 3 klaster sebagai berikut:



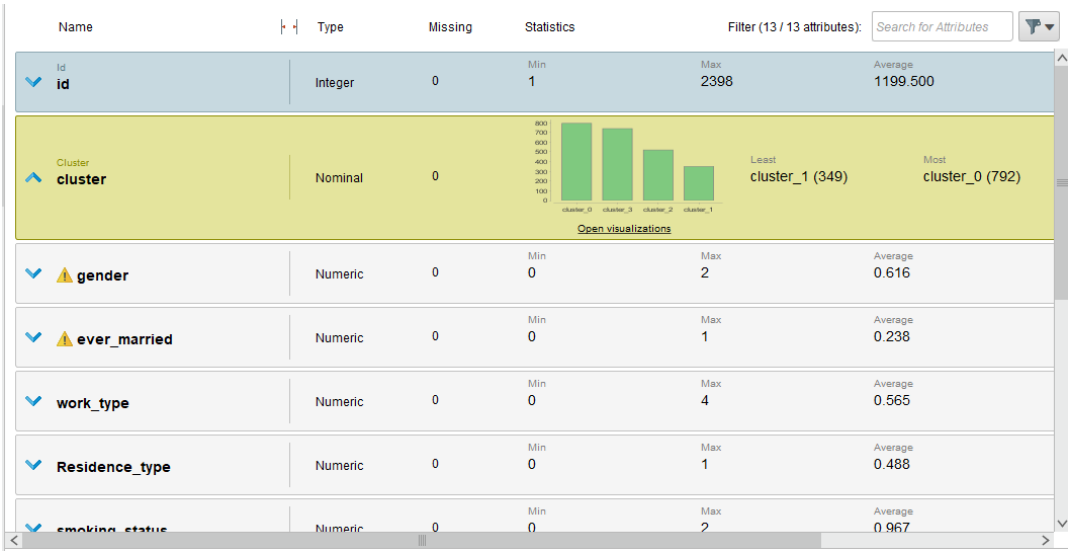
Gambar 4. 10 Visualisasi Hasil Klasterisasi

Klaster 1 memiliki presentase atau kemiripan data rendah sedangkan klaster 2 memiliki presentase atau kemiripan sedang dan klaster 3 memiliki presentase atau kemiripan yang banyak di antar klaster 1 dan 2. Berikut visualisasi hasil klaster yang terbentuk:

Row No.	id	cluster	gender	ever_married	work_type	Residence_t...	smoking_st...	age	hypertension	h
1	1	cluster_1	0	0	0	0	0	67	0	1
2	2	cluster_3	0	0	0	1	1	80	0	1
3	3	cluster_1	1	0	1	1	1	79	1	0
4	4	cluster_3	0	0	0	1	1	74	1	1
5	5	cluster_3	1	1	0	0	1	69	0	0
6	6	cluster_3	1	0	0	1	1	81	1	0
7	7	cluster_2	1	0	0	0	2	54	0	0
8	8	cluster_1	1	0	0	0	1	79	0	1
9	9	cluster_1	1	0	1	1	1	50	1	0
10	10	cluster_1	0	0	0	0	2	75	1	0
11	11	cluster_3	1	1	0	0	1	60	0	0
12	12	cluster_1	1	0	1	0	1	52	1	0
13	13	cluster_1	1	0	1	0	1	79	0	0
14	14	cluster_3	0	0	1	1	1	80	0	0

Gambar 4. 11 Visualisasi Hasil Cluster

Adapun secara statistik digambarkan sebagai berikut:



Gambar 4. 12 Statistik

Berdasarkan hasil klasterisasi, maka data set penyakit stroke dengan jumlah 3426 dibagi menjadi 4 karakteristik kemiripan data. Sebanyak 1434 memiliki karakteristik data yang sama dan termasuk dalam klaster 0. Sebanyak 1891

memiliki karakteristik data yang sama termasuk dalam klaster 1. Dan 501 memiliki karakteristik data yang sama dan termasuk pada klaster 2. Sedangkan 1284 data memiliki karakteristik yang mirip serta termasuk dalam klaster 3.

Pada tahap selanjutnya, kita akan melihat hasil klasterisasi yang dibagi menjadi 4 klaster, yaitu Klaster 0**,** Klaster 1**,** Klaster 2, dan Klaster 3. Setiap klaster mewakili grup individu dengan karakteristik gejala dan risiko stroke yang berbeda, berdasarkan atribut seperti Smoking status, heart disease, hypertension Glukosa, Bmi, dan Age.

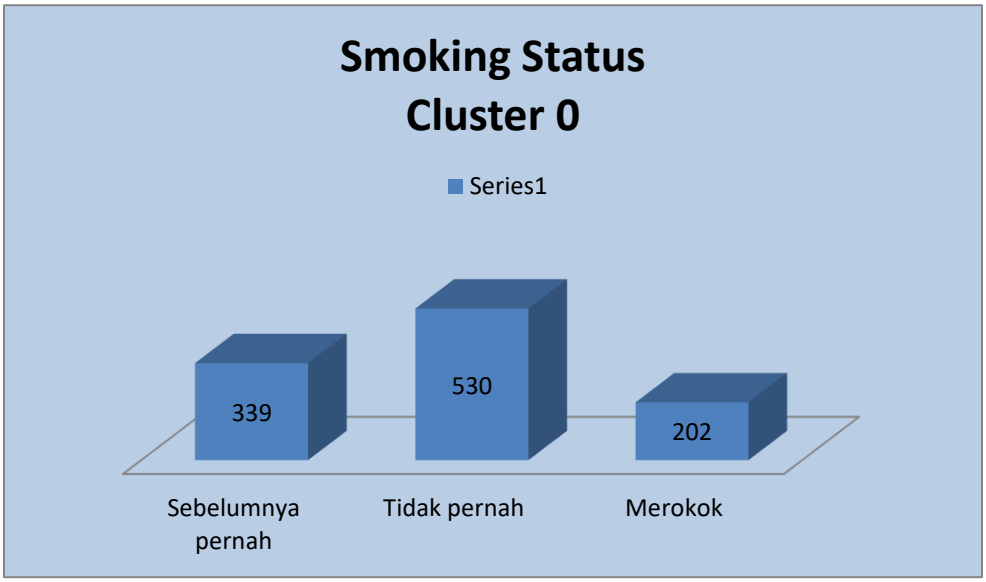
a. Cluster 0

   Klaster 0 merupakan kelompok individu dengan risiko tinggi terhadap stroke berdasarkan faktor-faktor kesehatan yang diamati. Dalam klaster ini, terdapat banyak individu yang memiliki riwayat penyakit jantung dan hipertensi, dua faktor utama yang berkontribusi terhadap risiko stroke. Selain itu, prevalensi obesitas (BMI tinggi) serta kebiasaan merokok yang cukup signifikan juga turut memperburuk kondisi kesehatan kelompok ini. Berikut visualisasi berdasarkan atribut nya :

- Smoking Status

  Keterangan :

  Sebanyak 339 orang memiliki riwayat merokok di masa lalu. Meskipun mereka telah berhenti, dampak dari kebiasaan merokok sebelumnya masih dapat berpengaruh terhadap kesehatan mereka, terutama dalam hal risiko penyakit kardiovaskular dan stroke. Sebanyak 530 orang tidak pernah merokok, yang menunjukkan bahwa kelompok ini memiliki risiko lebih rendah terkait dampak negatif dari kebiasaan merokok. Sebanyak 202 orang masih aktif merokok, yang berarti mereka memiliki risiko lebih tinggi terhadap berbagai masalah kesehatan, termasuk hipertensi, penyakit jantung, dan stroke.



Gambar 4. 13 Smoking Status Cluster 0

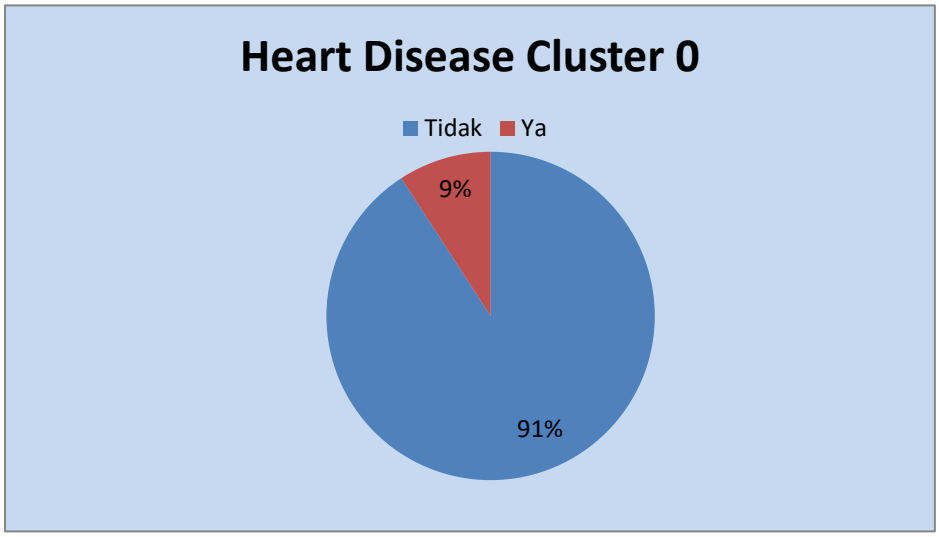
22,1% individu pernah merokok sebelumnya, yang berarti mereka memiliki paparan risiko akibat merokok di masa lalu.

34,5% individu tidak pernah merokok, yang menunjukkan bahwa sebagian besar kelompok ini tidak memiliki risiko dari kebiasaan merokok.

13,2% individu masih aktif merokok, yang dapat meningkatkan risiko penyakit kardiovaskular, termasuk stroke.

- Heart disease

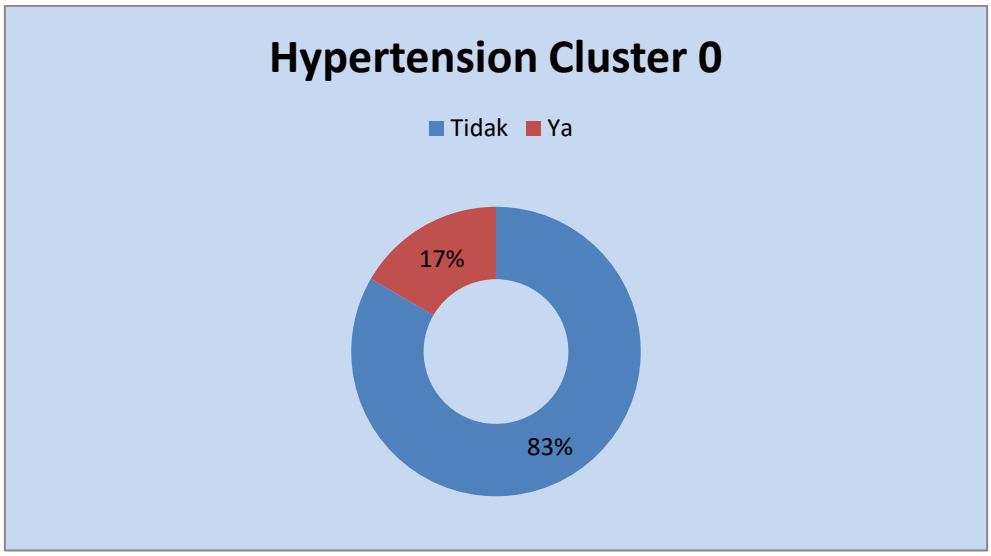
  Sebanyak 973 orang tidak memiliki riwayat penyakit jantung, yang menunjukkan bahwa mayoritas individu dalam klaster ini memiliki risiko yang lebih rendah terkait faktor ini. Sebanyak 98 orang memiliki penyakit jantung, yang berarti mereka memiliki risiko yang lebih tinggi untuk mengalami komplikasi kesehatan, termasuk stroke.



Gambar 4. 14 Heart Disease Cluster 0

- Hypertension

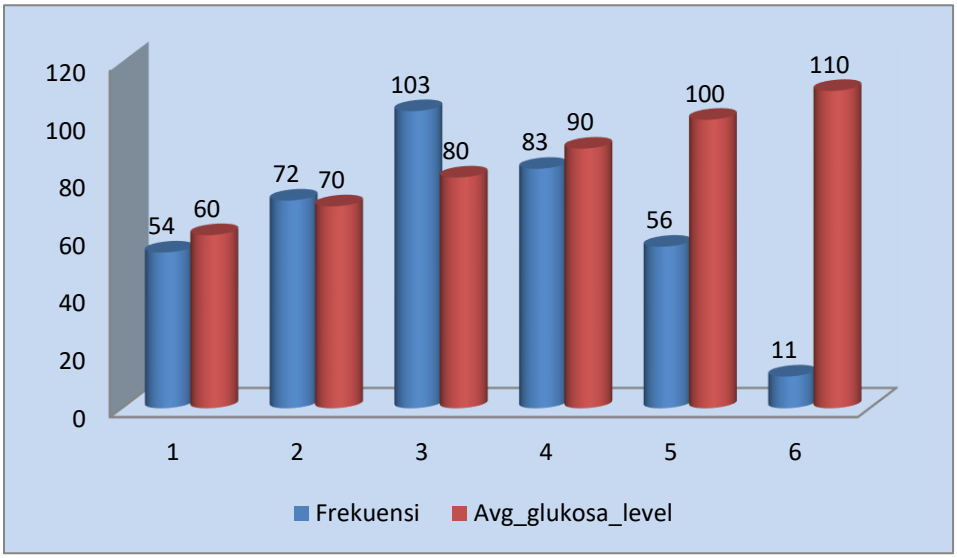
  Sebanyak 893 orang tidak memiliki riwayat hipertensi, yang menunjukkan bahwa mayoritas individu dalam klaster ini memiliki tekanan darah dalam batas normal. Sebanyak 178 orang memiliki hipertensi, yang berarti mereka berisiko lebih tinggi mengalami komplikasi kesehatan, termasuk stroke dan penyakit kardiovaskular lainnya.



Gambar 4. 15 Hypertension Cluster 0

- Avg Glukosa

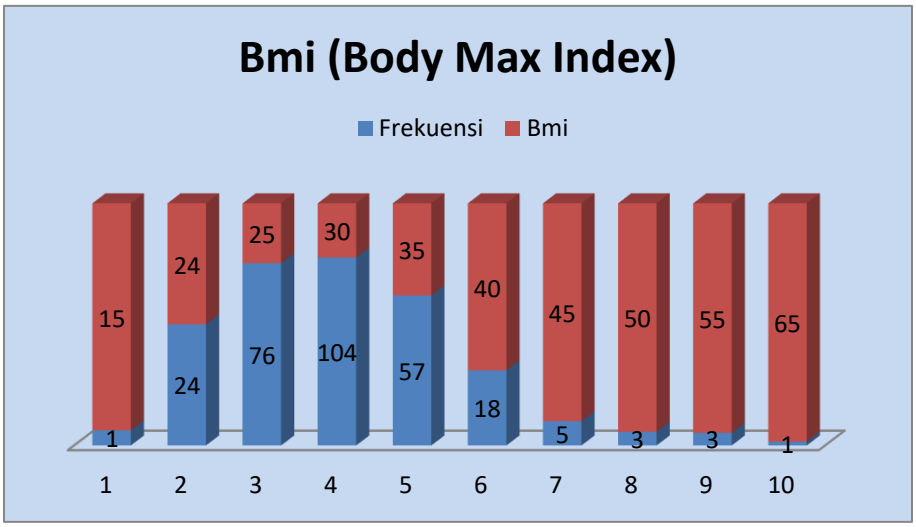
  Meskipun sebagian besar individu dalam Klaster 0 memiliki kadar glukosa dalam rentang normal, terdapat sejumlah individu yang memiliki kadar glukosa mendekati atau melebihi batas atas normal. Kadar glukosa yang tinggi dapat meningkatkan risiko berbagai komplikasi kesehatan, termasuk diabetes dan penyakit kardiovaskular, yang merupakan faktor utama dalam kejadian stroke.



Gambar 4. 16 Visualisasi Avg Glukosa Cluster 0

- Bmi

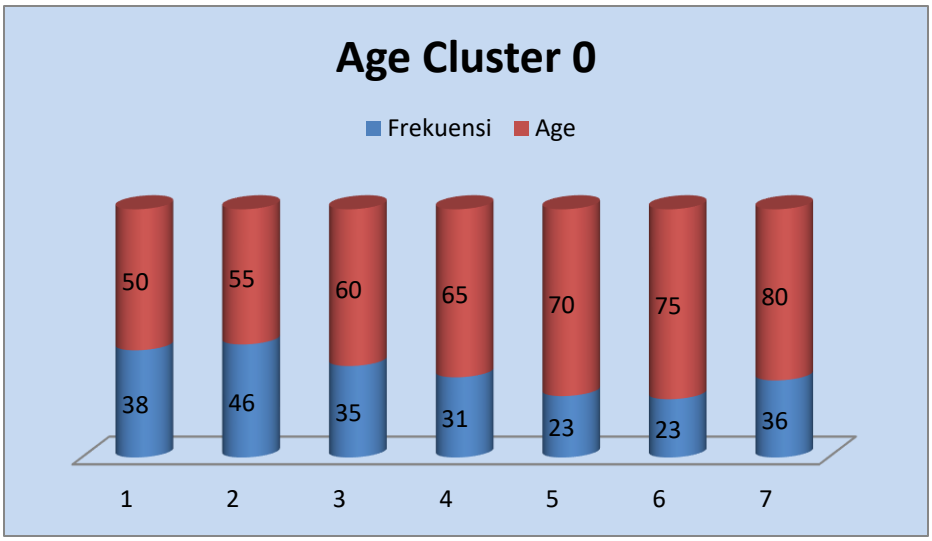
  Dalam Klaster 0, terdapat individu dengan BMI yang tinggi, yang dapat meningkatkan risiko berbagai masalah kesehatan, termasuk hipertensi, penyakit jantung, dan diabetes. Kondisi ini menjadi faktor utama yang dapat memicu stroke dan komplikasi kesehatan lainnya.



Gambar 4. 17 Visualisasi Bmi Cluster 0

- Age

  Klaster 0 didominasi oleh individu dengan usia yang lebih tua, yang secara alami memiliki risiko lebih tinggi terhadap berbagai penyakit kronis, termasuk hipertensi, diabetes, dan penyakit jantung. Faktor usia juga merupakan salah satu pemicu utama dalam meningkatnya kemungkinan seseorang mengalami stroke.



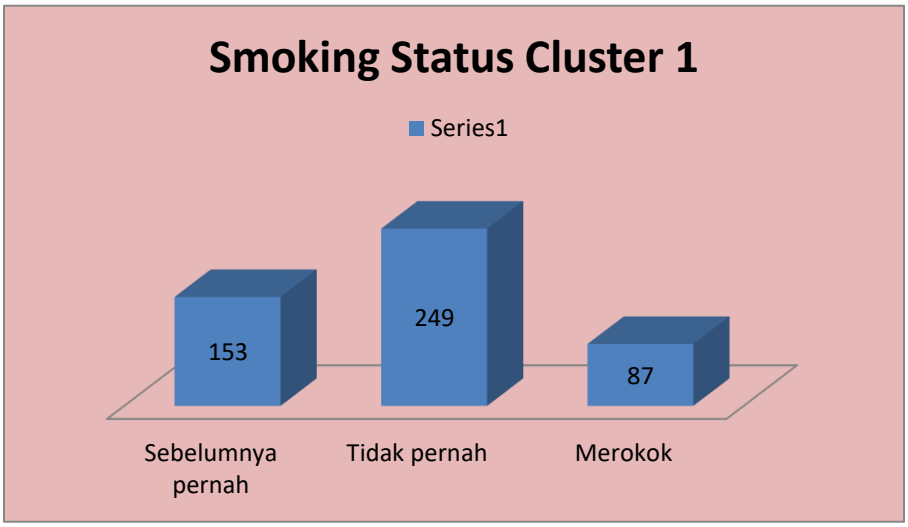
Gambar 4. 18 Age Cluster 0

b. Cluster 1

Klaster 1 juga termasuk dalam kelompok dengan risiko tinggi terhadap stroke. Hal ini disebabkan oleh tingginya jumlah individu dengan riwayat penyakit jantung dan hipertensi, yang merupakan faktor utama pemicu stroke. Selain itu, kadar glukosa darah yang sangat tinggi pada sebagian besar individu dalam klaster ini semakin meningkatkan potensi komplikasi kesehatan, terutama terkait dengan diabetes dan penyakit kardiovaskular.. Berikut visualisasi berdasarkan atribut nya :

- Smoking Status

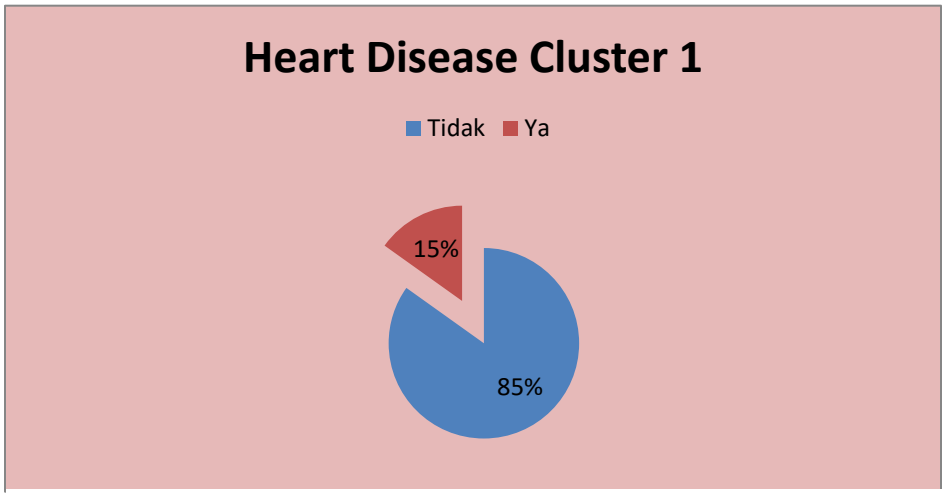
Klaster ini mencakup individu yang masih aktif merokok maupun yang pernah merokok sebelumnya. Kebiasaan merokok dapat menyebabkan kerusakan pembuluh darah, meningkatkan tekanan darah, dan mempercepat pembentukan plak di arteri, yang semuanya merupakan faktor risiko stroke.



Gambar 4. 19 Smoking Status Cluster 1

- Heart Disease

  Terdapat individu dalam klaster ini yang memiliki riwayat penyakit jantung. Kondisi ini meningkatkan risiko pembentukan gumpalan darah yang dapat menyumbat aliran darah ke otak dan menyebabkan stroke.

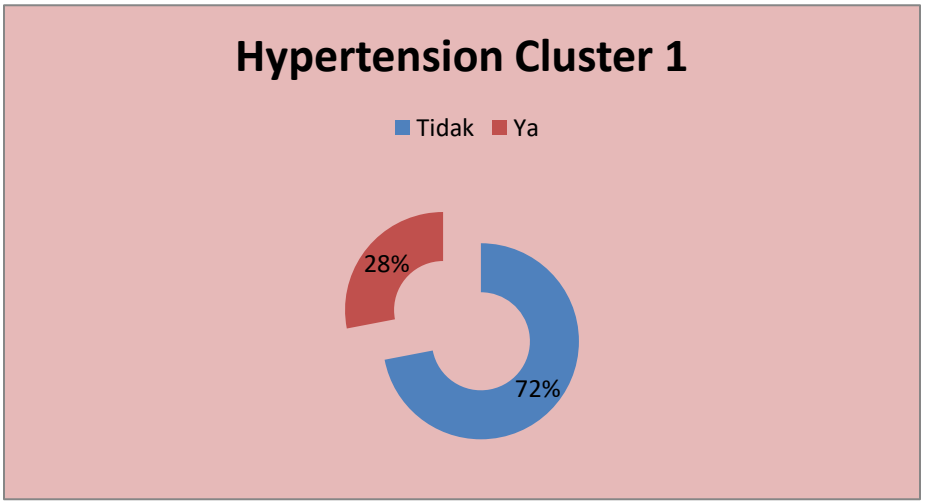


Gambar 4. 20 Heart Disease Cluster 1

- Hypertension

  Tekanan darah tinggi yang dialami oleh sebagian individu dalam klaster ini merupakan salah satu faktor utama dalam kejadian stroke. Hipertensi dapat

melemahkan dinding pembuluh darah dan meningkatkan risiko pecahnya pembuluh darah di otak.

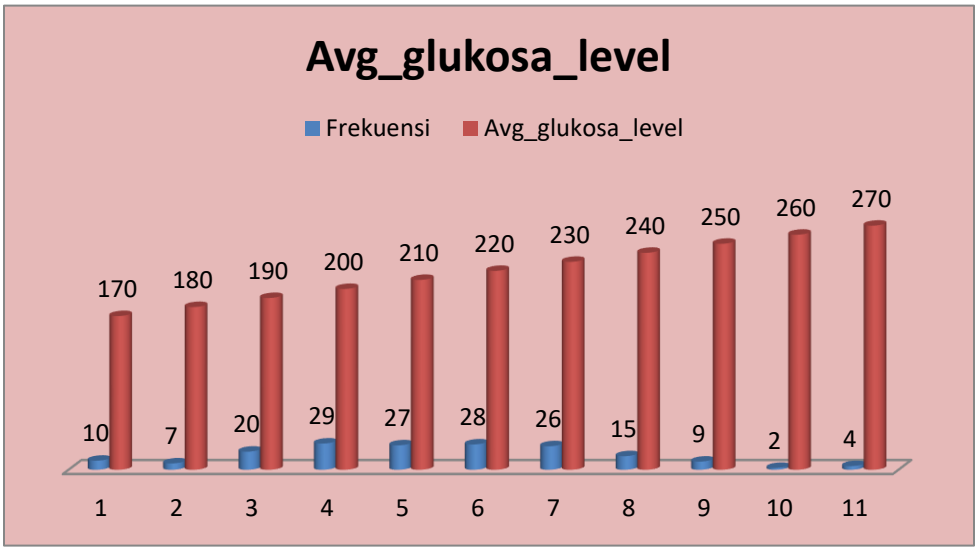


Gambar 4. 21 Hypertension Cluster 1

Sebanyak 7,62% individu dalam Klaster 1 menderita hipertensi, yang tetap menjadikan klaster ini masuk dalam kategori risiko sedang karena dampak hipertensi yang signifikan terhadap kemungkinan terjadinya stroke, terutama jika dikombinasikan dengan faktor risiko lainnya.

- Avg Glukosa

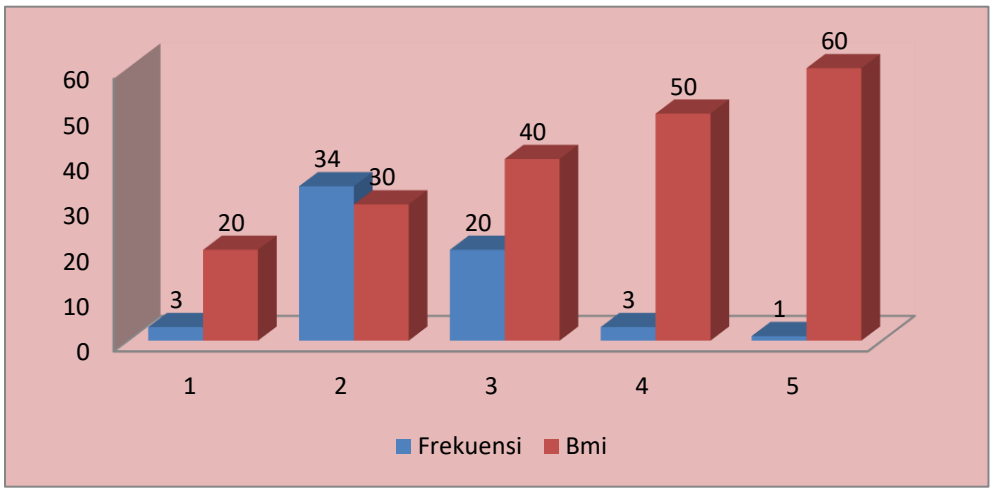
  Sebagian individu dalam klaster ini memiliki kadar glukosa yang tinggi, yang mengindikasikan risiko diabetes. Kadar glukosa yang tidak terkontrol dapat merusak pembuluh darah dan saraf, sehingga meningkatkan kemungkinan terjadinya stroke.



Gambar 4. 22 Avg Glukosa Cluster 1

- Bmi

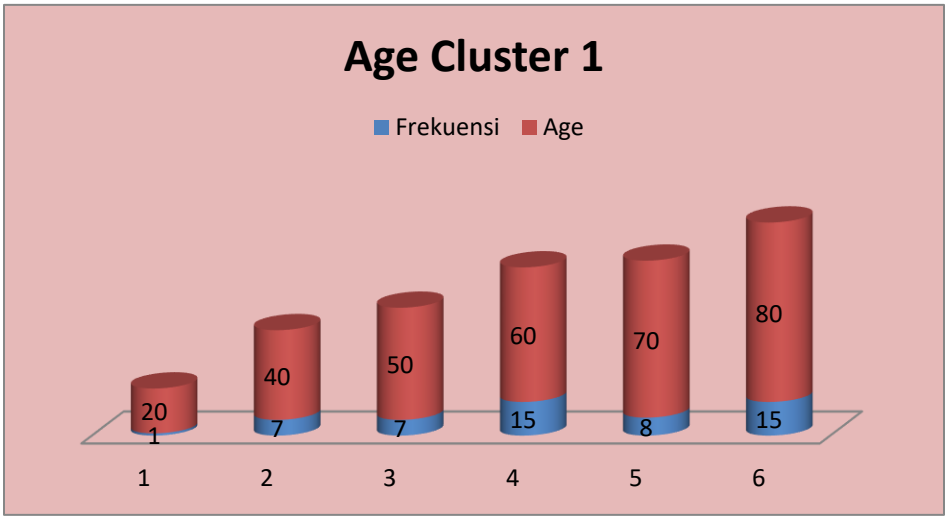
  16 Klaster ini mencakup individu dengan BMI tinggi, yang menandakan adanya risiko obesitas. Obesitas sering dikaitkan dengan peningkatan tekanan darah, resistensi insulin, dan gangguan metabolisme yang dapat memperburuk faktor risiko stroke.



Gambar 4. 23 Bmi Cluster 1

- Age

  Individu dalam klaster ini sebagian besar berada pada usia lanjut, yang secara alami memiliki risiko lebih tinggi terhadap penyakit degeneratif seperti hipertensi, diabetes, dan penyakit kardiovaskular.



Gambar 4. 24 Age Cluster 1

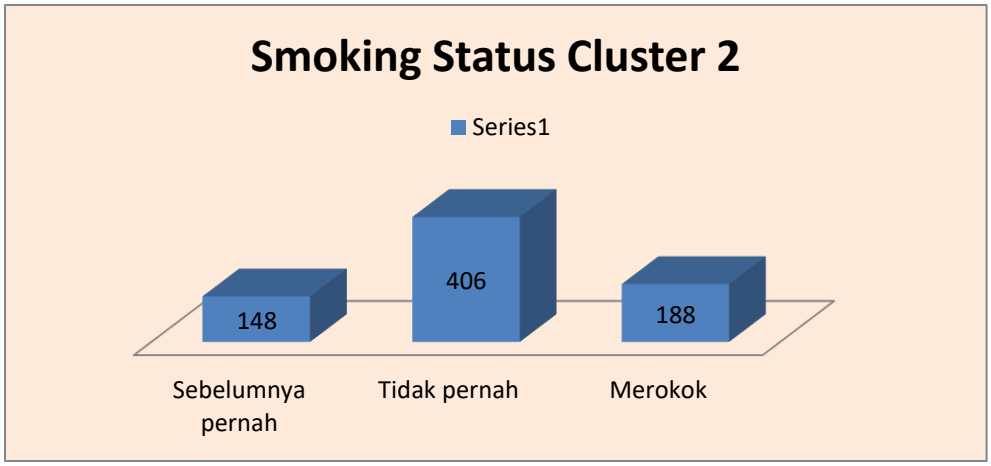
c. Cluster 2

Klaster 2 dikategorikan sebagai kelompok dengan risiko sedang terhadap stroke. Dalam klaster ini, jumlah individu dengan hipertensi dan penyakit jantung lebih rendah dibandingkan dengan klaster berisiko tinggi. Namun, terdapat sejumlah individu dengan kadar glukosa darah yang cukup tinggi, yang dapat menjadi faktor pemicu komplikasi kesehatan jika tidak dikelola dengan baik. Berikut Visualisasi berdasarkan atribut :

- Smoking Status

  Keterangan :

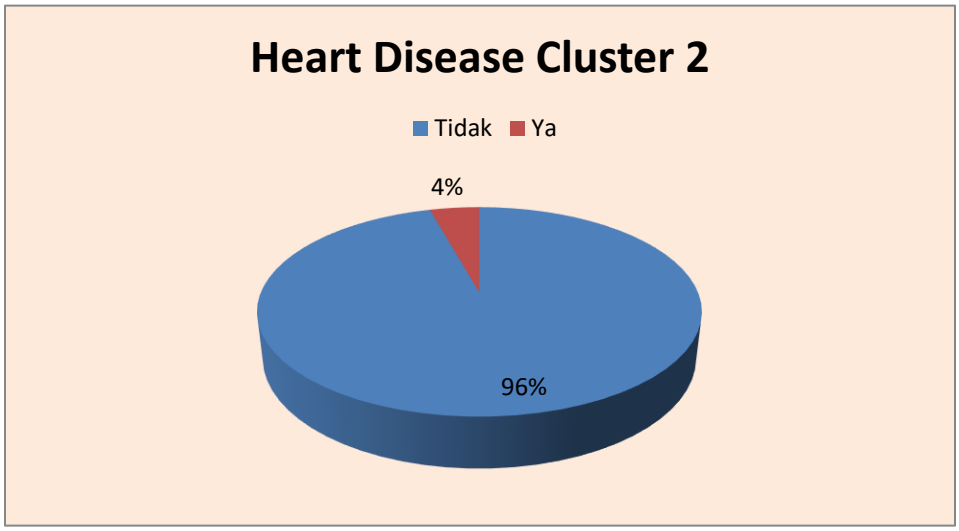
  Klaster ini mencakup individu yang merokok maupun yang pernah merokok sebelumnya. Meskipun jumlah perokok tidak sebanyak klaster berisiko tinggi, paparan zat berbahaya dari rokok tetap dapat memengaruhi kesehatan jantung dan pembuluh darah dalam jangka panjang.



Gambar 4. 25 Smoking Status Cluster 2

- Heart Disease

  Beberapa individu dalam klaster ini memiliki riwayat penyakit jantung, tetapi dalam jumlah yang lebih sedikit dibandingkan klaster berisiko tinggi. Namun, tetap ada potensi komplikasi kardiovaskular yang dapat berkembang menjadi stroke, terutama jika tidak ditangani dengan baik.

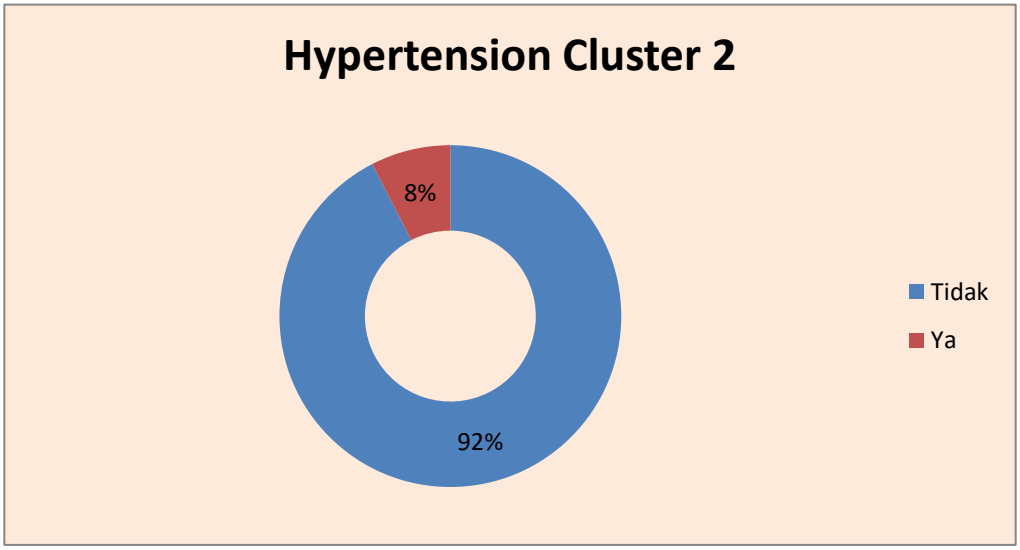


Gambar 4. 26 Heart Disease Cluster 2

- Hypertension

  Hipertensi juga ditemukan dalam klaster ini, meskipun jumlahnya tidak terlalu dominan. Tekanan darah yang sedikit meningkat tetap bisa

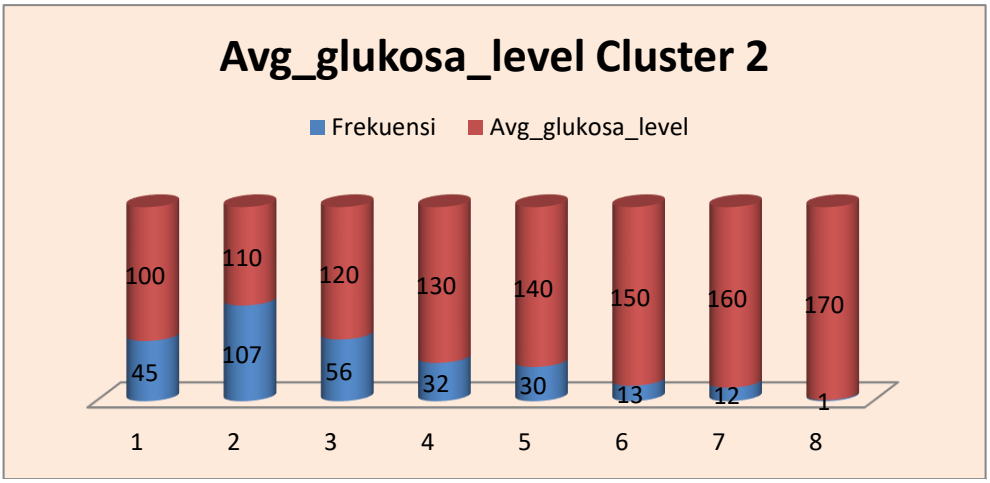
memberikan dampak negatif terhadap kesehatan pembuluh darah dan meningkatkan risiko stroke dalam jangka panjang.



Gambar 4. 27 Hypertension Cluster 2

- Avg Glukosa

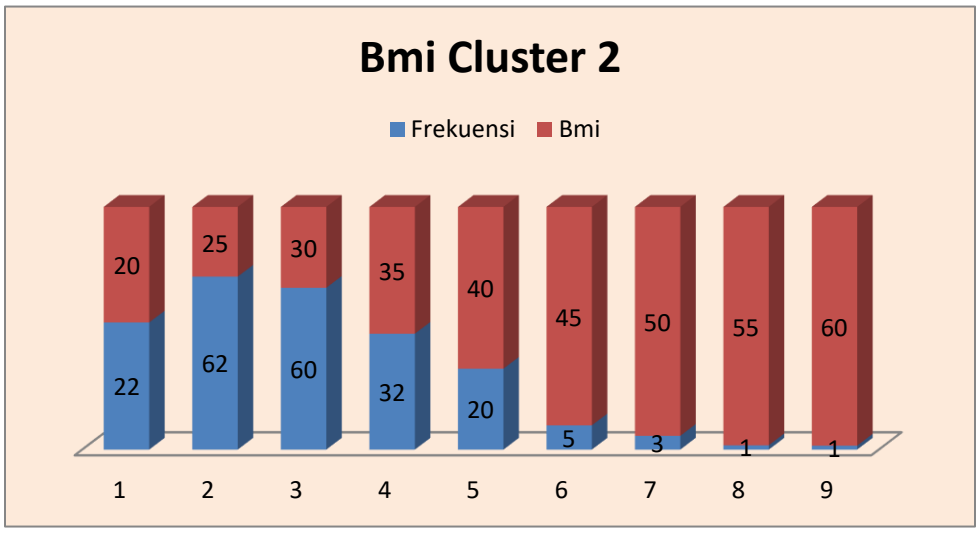
  Klaster ini menunjukkan adanya individu dengan kadar glukosa yang mulai meningkat, meskipun tidak setinggi kelompok berisiko tinggi. Peningkatan kadar glukosa yang terus-menerus dapat menjadi awal dari gangguan metabolik, termasuk diabetes, yang jika tidak dikontrol dapat meningkatkan risiko stroke.



Gambar 4. 28 Avg Glukosa Cluster 2

- Bmi

  Sebagian individu dalam klaster ini memiliki BMI yang lebih tinggi dari normal, tetapi tidak sebanyak pada klaster risiko tinggi. Kelebihan berat badan tetap dapat berkontribusi pada hipertensi dan resistensi insulin, yang dalam jangka panjang dapat meningkatkan risiko stroke.

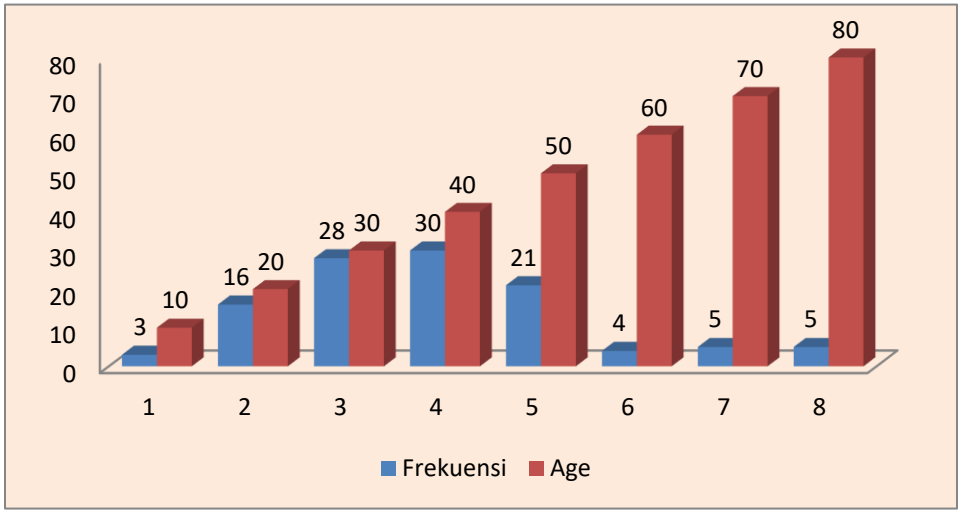


Gambar 4. 29 Bmi Cluster 2

- Age

  Klaster ini mencakup individu dari berbagai kelompok usia, termasuk

beberapa individu yang lebih muda. Meskipun usia bukan faktor risiko dominan dalam klaster ini, tetap diperlukan perhatian khusus bagi individu yang memiliki faktor risiko lain.



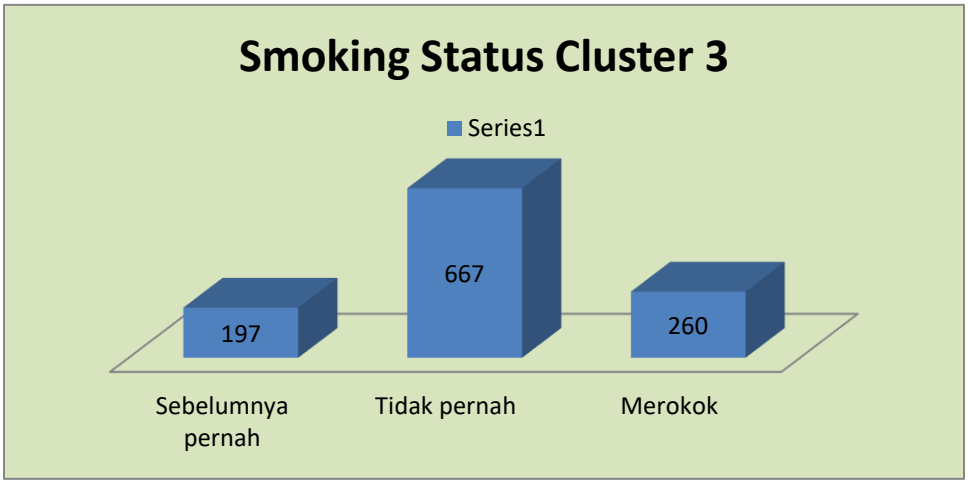
Gambar 4. 30 Age Cluster 2

d. Cluster 3

Klaster 3 merupakan kelompok dengan risiko rendah terhadap stroke. Hal ini didukung oleh tidak adanya individu dengan riwayat penyakit jantung, serta jumlah penderita hipertensi yang sangat sedikit dibandingkan dengan klaster lainnya. Selain itu, mayoritas individu dalam klaster ini memiliki kadar glukosa darah yang normal dan indeks massa tubuh yang lebih terkontrol. Berikut visualisasi nya berdasarkan atribut nya:

- Smoking Status

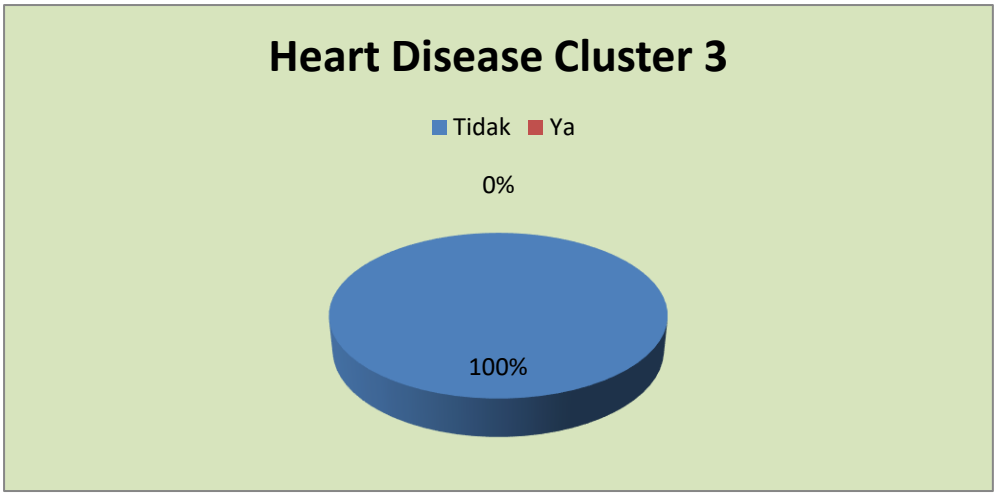
  Sebagian besar individu dalam klaster ini tidak merokok atau belum pernah merokok, sehingga risiko yang berkaitan dengan paparan zat berbahaya dari rokok terhadap sistem kardiovaskular lebih kecil dibandingkan dengan klaster lainnya.



Gambar 4. 31 Smoking Status Cluster 3

- Heart Disease

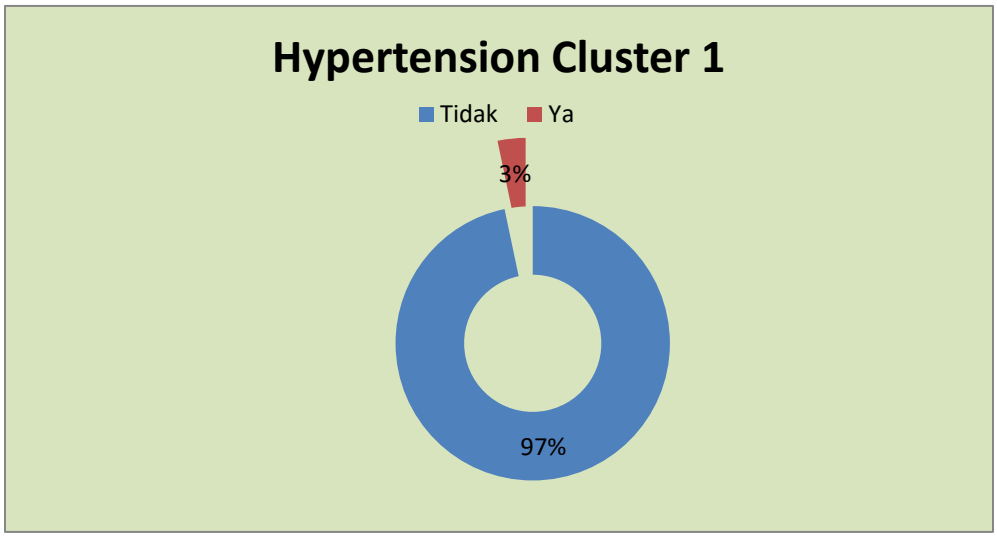
  Tidak ada individu dalam klaster ini yang memiliki riwayat penyakit jantung, yang menunjukkan bahwa faktor risiko utama untuk komplikasi kardiovaskular sangat rendah.



Gambar 4. 32 Heart Disease Cluster 3

- Hypertension

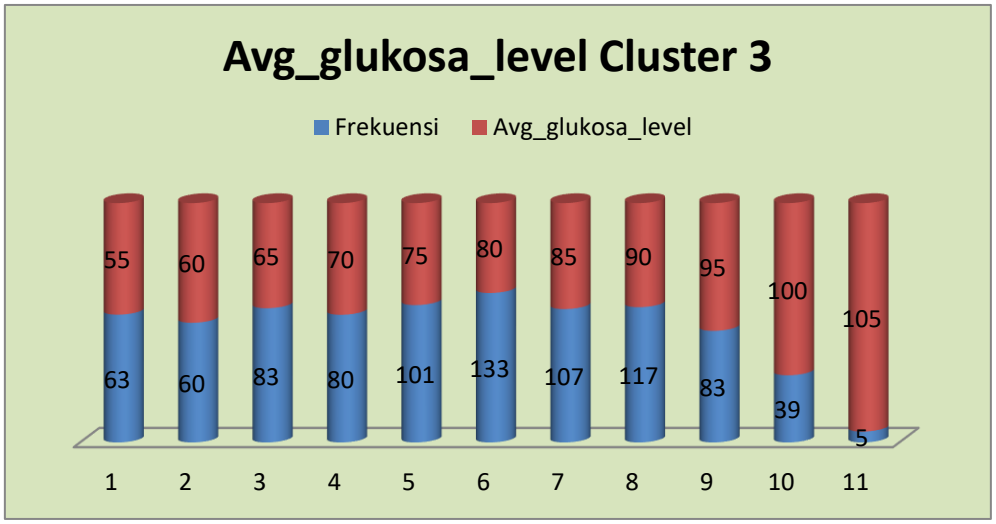
  Mayoritas individu dalam klaster ini tidak mengalami hipertensi, sehingga tekanan darah mereka relatif stabil, mengurangi kemungkinan kerusakan pada pembuluh darah yang bisa menyebabkan stroke.



Gambar 4. 33 Hypertension Cluster 3

- Avg Glukosa

  Kadar glukosa dalam klaster ini berada dalam rentang normal, yang menandakan risiko rendah terhadap diabetes dan gangguan metabolik lainnya yang dapat berkontribusi terhadap stroke.

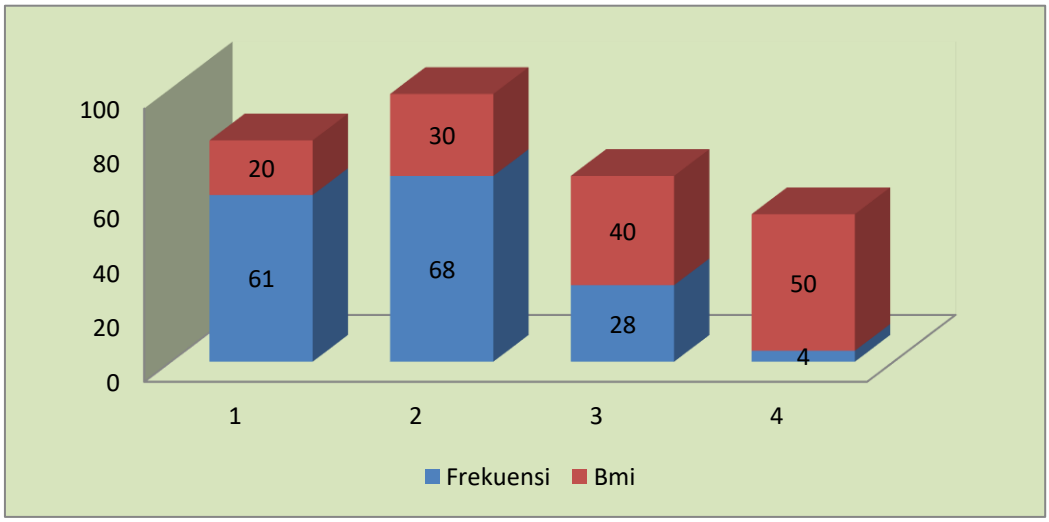


Gambar 4. 34 Avg Glukosa Cluster 3

- Bmi

  Sebagian besar individu dalam klaster ini memiliki BMI yang berada dalam rentang normal, yang mengindikasikan keseimbangan berat badan yang lebih

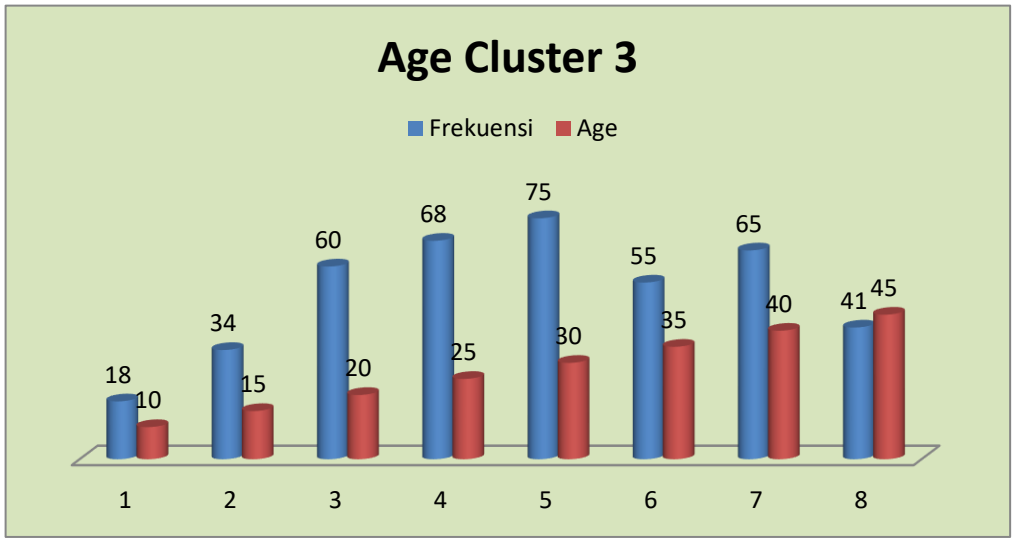
baik serta risiko yang lebih rendah terhadap penyakit terkait obesitas seperti hipertensi dan diabetes.



Gambar 4. 35 Bmi Cluster 3

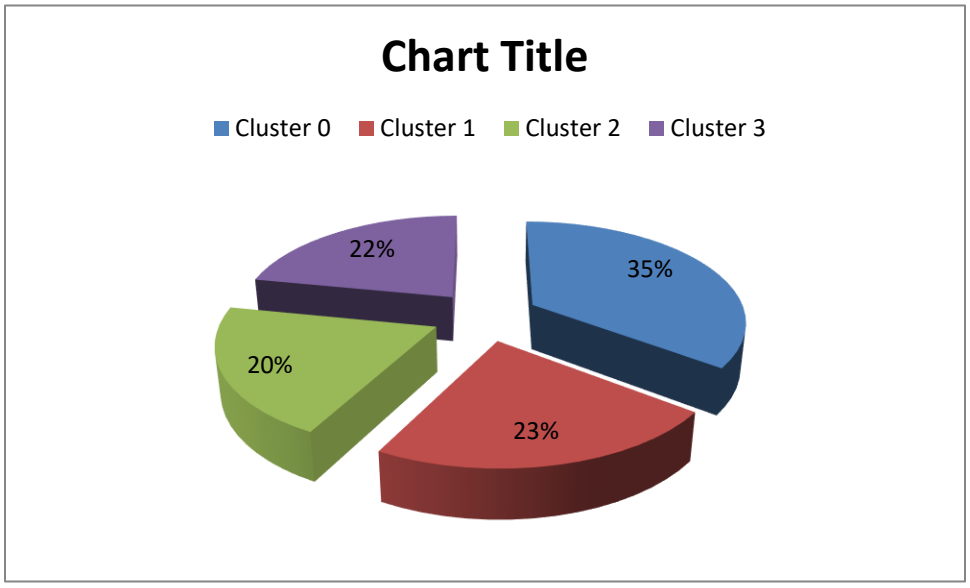
- Age

  Individu dalam klaster ini cenderung lebih muda dibandingkan klaster lainnya, yang secara alami mengurangi risiko stroke karena sistem kardiovaskular masih dalam kondisi lebih baik dan lebih mampu beradaptasi terhadap perubahan kesehatan.



Gambar 4. 36 Age Cluster 3

- Visualisasi Keseluruhan



Gambar 4.37 Visua1isasi Keseluruhan

Secara keseluruhan, pembagian klaster menunjukkan bahwa Klaster 0 merupakan klaster dengan proporsi tertinggi, mencakup 35% dari total data. Klaster 1 berada di posisi kedua dengan proporsi 23%, menunjukkan tingkat yang cukup tinggi namun masih di bawah Klaster 0. Sementara itu, Klaster 2 menjadi klaster dengan proporsi terendah, yaitu 20%, dan Klaster 3 menempati posisi sedikit lebih tinggi dari Klaster 2 dengan proporsi 22%.