Journal of Science and Social Research ISSN 2615 – 4307 (Print)
Oct 2026, IX (5): 8247 – 8257 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

ANALISIS POLA PERSEBARAN HIPERTENSI DI SUMATERA UTARA DENGAN MENGGUNAKAN ALGORITMA K-MEANS

Eldo Yusri[1], Azrai Sirait[2]
Universitas Asahan
e-mail: [1]eldoyusri9@gmail.com, [2]azraimu3@gmail.com

Abstract: *The fluctuating spread of hypertension requires accurate mapping for the allocation of medical resources. Often, regions with a consistent level of cases escape government attention compared to regions experiencing the highest case spikes in a specific year. Addressing this issue, this study aims to cluster 33 regencies/cities in North Sumatra using the K-Means Clustering algorithm to map the stability of disease spread based on 5-year historical data 2020–2024. The features used include mean prevalence X_1, growth rate X_2, and fluctuation level X_3. Data normalization was carried out using Min-Max Normalization and measured using Euclidean Distance. System testing reached convergence at the 3rd iteration, grouping regions into three clusters: High Priority C1; 15 regions, Medium Priority C2; 9 regions, and Low Priority C3; 9 regions. Quality evaluation using Silhouette Score resulted in a global average score of +0.3941 Weak Structure. The developed web-based system serves as a decision support tool for health authorities in formulating targeted medical intervention strategies.*

Keyword: *K-Means Clustering, Hypertension, Euclidean Distance, Min-Max Normalization, Silhouette Score..*

Abstrak: Penyebaran penyakit hipertensi yang berfluktuasi memerlukan pemetaan akurat dalam penentuan alokasi sumber daya medis. Selama ini, wilayah yang secara konsisten berada pada tingkat kasus tertentu sering kali luput dari perhatian pemerintah dibandingkan wilayah yang mengalami lonjakan tertinggi pada satu waktu tertentu. Penelitian ini bertujuan mengelompokkan 33 kabupaten/kota di Sumatera Utara menggunakan algoritma K-Means Clustering untuk memetakan kestabilan penyebaran penyakit hipertensi berdasarkan data historis 5 tahun 2020–2024. Fitur yang digunakan meliputi rata-rata prevalensi X_1, laju pertumbuhan X_2, dan tingkat fluktuasi X_3. Seluruh data dinormalisasi dengan metode Min-Max Normalization dan dihitung jaraknya menggunakan Euclidean Distance. Hasil pengujian sistem mencapai titik konvergen pada iterasi ke-3, mengelompokkan wilayah ke dalam tiga klaster: Prioritas Tinggi C1; 15 wilayah, Prioritas Sedang C2; 9 wilayah, dan Prioritas Rendah C3; 9 wilayah. Evaluasi kualitas struktur klaster menggunakan Silhouette Score menghasilkan skor rata-rata global sebesar +0,3941 kategori Weak Structure. Sistem berbasis web yang dibangun dapat membantu instansi kesehatan dalam penentuan strategi penanganan kesehatan yang objektif dan tepat sasaran.

Kata kunci: K-Means Clustering, Hipertensi, Euclidean Distance, Min-Max Normalization, Silhouette Score

PENDAHULUAN

Setiap tahun, Dinas Kesehatan Provinsi Sumatera Utara mempublikasikan laporan Profil Kesehatan sebagai sumber informasi kondisi kesehatan masyarakat. Dalam laporan tersebut, hipertensi tercatat sebagai salah satu penyakit tidak menular yang dominan di berbagai kabupaten/kota, dengan variasi capaian pelayanan antarwilayah, seperti

Kabupaten Padang Lawas yang persentase kasusnya relatif lebih tinggi dibandingkan daerah lain.

Namun, data tersebut umumnya hanya disajikan dalam bentuk grafik batang statis yang menampilkan capaian kasus per tahun secara terpisah, tanpa disertai analisis lanjutan yang mampu menghubungkan pola antarwaktu. Penyajian visual semacam ini bersifat sekilas dan deskriptif, sehingga cenderung hanya menonjolkan wilayah dengan nilai tertinggi pada tahun tertentu, tanpa mengungkap keterkaitan tren maupun pola perubahan kasus dari tahun ke tahun. Akibatnya, karakteristik dan pola persebaran hipertensi antarwilayah belum dapat dijelaskan secara menyeluruh, dan pola tersembunyi di balik lonjakan atau penurunan kasus antarwaktu sulit untuk ditangkap hanya dengan mengamati grafik secara visual.

Padahal, data historis yang tersedia mencakup rentang waktu yang cukup panjang, yaitu lima tahun 2020 hingga 2024, dan berpotensi besar untuk ditambang data mining guna mengidentifikasi pola pengelompokan wilayah berdasarkan kemiripan karakteristiknya, seperti nilai rata-rata, laju pertumbuhan, dan tingkat fluktuasi kasus. Perlu ditegaskan bahwa penelitian ini murni berfokus pada pola pengelompokan geografis kabupaten/kota, bukan pada penyebab medis, faktor klinis, maupun etiologi hipertensi.

Tanpa adanya proses analisis komputasi yang cerdas, data tersebut hanya berfungsi sebagai laporan statistik pasif yang belum optimal mendukung pengambilan keputusan, sehingga menyulitkan instansi kesehatan dalam menentukan zonasi wilayah prioritas penanganan secara objektif. Oleh karena itu, diperlukan suatu metode Machine Learning yang mampu mengolah data secara sistematis, salah satunya algoritma K-Means, yang mengelompokkan wilayah berdasarkan tingkat kemiripan karakteristiknya secara otomatis.

Selain itu, penelitian ini juga ingin melihat kestabilan posisi setiap wilayah dari tahun ke tahun, apakah suatu wilayah mengalami kenaikan kasus secara konsisten, berfluktuasi, atau cenderung stabil pada kelompok yang sama. Hal ini penting karena wilayah yang secara konsisten berada pada tingkat kasus tertentu, meski tidak selalu menempati peringkat teratas, justru cenderung luput dari perhatian dibanding wilayah dengan kasus tertinggi pada satu tahun tertentu. Padahal, konsistensi tersebut dapat menjadi indikator penting dalam menentukan prioritas penanganan, karena bukan hanya besarnya angka pada satu waktu yang perlu diperhatikan, melainkan juga pola keteraturan persebaran kasus dalam jangka panjang.

Penerapan algoritma K-Means dalam bidang kesehatan telah banyak dilakukan, namun masih memiliki keterbatasan. Penelitian oleh Meida et al. (2023) menerapkan K-Means untuk mengelompokkan tingkat kerawanan kabupaten/kota di Sumatera Utara, namun hanya berfokus pada penyakit menular seperti Tuberkulosis dan Demam Berdarah. Penelitian lain oleh Rhodiyah (2025) menggunakan algoritma K-Means untuk menganalisis pola penyakit kronis, termasuk hipertensi, tetapi cakupan wilayahnya masih terbatas pada satu puskesmas di tingkat kelurahan sehingga belum merepresentasikan kondisi pada skala makro. Selanjutnya, penelitian oleh Fani et al. (2021) memanfaatkan K-Means untuk pemetaan persebaran wilayah terdampak di Provinsi Sumatera Utara, namun objek yang diteliti adalah penyebaran Covid-19 yang memiliki karakteristik berbeda dengan penyakit tidak menular. Selain itu, Laelatul et al. (2025) menerapkan metode K-Means untuk memetakan persebaran penyakit pada tingkat kecamatan, namun analisis yang dilakukan masih bersifat statis dan belum mengkaji pola tren perubahan data antarwaktu secara spesifik.

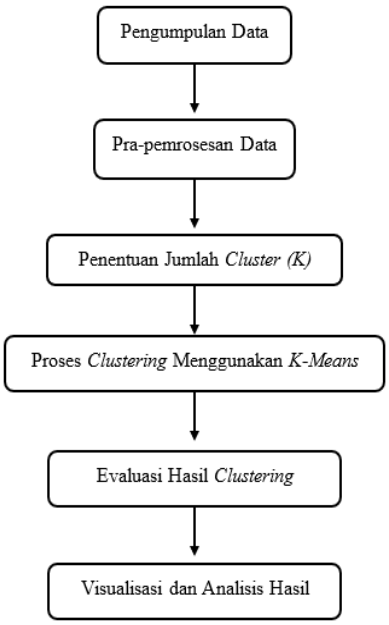
Berdasarkan tinjauan terhadap penelitian-penelitian tersebut, dapat disimpulkan bahwa penggunaan algoritma K-Means dalam bidang kesehatan masih memiliki keterbatasan pada aspek

pemanfaatan data historis dinamis. Penelitian sebelumnya umumnya menggunakan cakupan wilayah yang terbatas dan belum mengoptimalkan data multi-tahun untuk menganalisis pola perubahan persebaran penyakit secara mendalam dan dinamis.

Berbeda dari penelitian-penelitian sebelumnya, penelitian ini bertujuan untuk menganalisis pola persebaran hipertensi menggunakan data historis selama lima tahun 2020–2024 pada seluruh 33 kabupaten/kota di Provinsi Sumatera Utara. Analisis dilakukan dengan algoritma K-Means untuk mengelompokkan wilayah ke dalam kategori tinggi, sedang, dan rendah secara objektif berdasarkan tren data tersebut. Dengan demikian, keterbaruan penelitian ini terletak pada pemanfaatan rekayasa fitur dari data historis multi-tahun pada skala provinsi untuk mengidentifikasi pola persebaran hipertensi secara dinamis, serta cakupan penelitian pada tingkat provinsi yang melibatkan seluruh wilayah secara menyeluruh. Hasil penelitian ini diharapkan dapat menjadi sistem pendukung keputusan yang membantu instansi kesehatan dalam memberikan gambaran prioritas penanganan secara lebih komprehensif, otomatis, dan tepat sasaran.

METODE

Metode penelitian yang digunakan dalam penelitian ini adalah metode data mining dengan pendekatan clustering menggunakan algoritma K-Means. Metode ini dipilih karena kemampuannya dalam mengelompokkan data secara cepat, efisien untuk dataset berskala besar, serta kemudahannya dalam menginterpretasikan hasil klaster. Tahapan penelitian dilakukan secara sistematis pada gambar 1.



Gambar 1. Kerangka Kerja Penelitian

Tahapan penelitian dilakukan secara sistematis yang meliputi: (1) Pengumpulan Data, yaitu menghimpun data hipertensi Kabupaten/Kota di Sumatera Utara dari sumber resmi Kementerian Kesehatan Provinsi Sumatera Utara. (2) Pra-pemrosesan Data, yaitu membersihkan data dari nilai kosong, ganda, atau tidak sesuai, serta mentransformasikannya ke dalam tiga indikator (rata-rata prevalensi, laju pertumbuhan, dan tingkat fluktuasi). (3) Penentuan Jumlah Cluster (K), yaitu menentukan nilai K optimal menggunakan metode Elbow berdasarkan grafik Sum of Squared Error (SSE). (4) Proses Clustering Menggunakan K-Means, yaitu mengelompokkan data ke dalam kategori tinggi, sedang, dan rendah berdasarkan jarak Euclidean terkecil ke centroid yang diperbarui secara berulang. (5) Evaluasi Hasil Clustering, yaitu menguji kualitas dan validitas hasil pengelompokan menggunakan metode Silhouette Score; serta (6) Visualisasi dan Analisis Hasil, yaitu menyajikan distribusi cluster secara visual dan menganalisis karakteristik pola

persebaran hipertensi sebagai bahan pertimbangan pengambilan keputusan di bidang kesehatan.

Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan menghimpun data kasus hipertensi di wilayah Sumatera Utara dari tahun 2020 hingga 2024 yang mencakup 33 kabupaten dan kota dari sumber resmi Kementerian Kesehatan Provinsi Sumatera Utara, di mana pengumpulan data yang akurat dan relevan sangat penting agar proses analisis yang dilakukan valid serta dapat dipertanggungjawabkan. Data tersebut kemudian diolah ke dalam tiga variabel utama, yaitu variabel X_1 (rata-rata) yang diperoleh dari jumlah keseluruhan kasus hipertensi tahun 2020–2024 dibagi dengan lima untuk merepresentasikan besaran atau beban kasus secara umum, variabel X_2 (laju pertumbuhan) yang dihitung dari selisih data tahun 2024 dengan tahun 2020 dibagi tahun 2020 lalu dikali 100% untuk merepresentasikan arah tren perubahan kasus, serta variabel X_3 (fluktuasi) yang diperoleh dari standar deviasi deret angka selama lima tahun untuk merepresentasikan tingkat kestabilan dari tahun ke tahun.

No	Variabel	Indikator	Deskripsi
1	X1	Rata-rata Kasus	Merepresentasikan besaran atau beban kasus hipertensi secara umum.
2	X2	Laju	Merepresentasikan arah tren perubahan kasus dari tahun awal ke tahun akhir.
3	X3	Fluktuasi	Merepresentasikan tingkat kestabilan atau variabilitas kasus dari tahun ke tahun.

Pra-pemrosesan Data

Setelah data terkumpul tahap selanjutnya adalah Pemrosesan Data (Prepro¬cessing Data). Dimana proses awal yaitu permbersihan data agar dapat di analisis dengan sangat baik. Permbersihan ini dilakukan untuk mengatasi data kosong, data ganda dan data yang tidak sesuai. Selain itu, data di tranformasi menjadi 3 indikator yaitu rata-rata pravalensi, laju pertumbuhan dan tingkat fluktuasi.

Penentuan Jumlah Cluster (K)

Setelah data melalui tahap pra-pemrosesan, langkah selanjutnya adalah menentukan jumlah cluster (K) yang akan digunakan dalam proses clustering. Penentuan jumlah cluster merupakan tahap penting karena akan mempengaruhi hasil pengelompokan data. Dalam penelitian ini, penentuan nilai K dilakukan menggunakan metode Elbow, yaitu dengan melihat grafik hubungan antara jumlah cluster dan nilai sum of squared error (SSE). Titik siku (elbow) pada grafik menunjukkan jumlah cluster yang optimal, yaitu ketika penambahan jumlah cluster tidak lagi memberikan penurunan signifikan terhadap nilai SSE. Dengan menggunakan metode ini, jumlah cluster yang dipilih diharapkan mampu merepresentasikan pola persebaran hipertensi secara optimal.
Proses Clustering Menggunakan K-Means
Pada tahap ini dilakukan proses utama yaitu pengelompokkan data menggunakan algoritma K-Means. Data hipertensi yang sudah melalui proses Pra-pemrosesan akan di kelompokkan sesuai dengan kemiripan dengan metode K-Means. Proses ini dimulai dari menentukan jumlah cluster yang diinginkan, kemudian menentukan centroid secara acak. Selanjutnya dilakukan perhitungan jarak dengan centroid dengan menggunakan metode

Euclidean Distance. Kemudian data akan dimasukkan kedalam cluster dengan jarak terdekat. Setelah itu centroid diperbarui berdasarkan rata rata dalam masing-masing cluster. Proses ini
dilakukan secara berulang sampai centroid tidak mengalami perubahan secara signifikan. Melalui proses ini data hipertensi dalam dikelompokkan kedalam beberapa kategori yaitu tinggi, sedang dan rendah sehingga pola persebaran dapat di ketahui dengan jelas.

Evaluasi Hasil Clustering

Setelah proses clustering selesai, di perlukan porses evaluasi hasil untuk menguji kualitas hasil clustering yang dilakukan. Evaluasi dilakukan dengam metode Silhouette Score untuk mengukur kualitas dari tingkat cluster berdasarkan kedekatan antar data dalam cluster dan jarak dari cluster yang berbeda. Tahap ini sangat penting karena dapat memastikan bahwa cluster benar benar menceriminkan pola yang terdapat pada data dan dapat digunakan sebagai dasar dalam pengambilan keputusan.

Visualisasi dan Analisis Hasil

Tahapan terakhir adalah Visualisasi dan Analisis Hasil. Pada tahapan ini hasil dari pengelompokan ditampilan kedalam visualisasi agar lebih mudah dipahami. Visualisasi ini berguna untuk memperlihatkan distribusi cluster sertam membantu mengindentifikasi pola persebaran hipertensi. Selanjutnya dilakukan analisis terhadap hasil clustering untuk mengetahui karakteristik masing masing cluster dan pola persebaran terbentuk, Hasil dari analsisi ini diharapkan bisa memberikan informasi yang bermanfaat keapda terkait, khususnya dalam bidang kesehatan.

HASIL DAN PEMBAHASAN

Analisa data

Berdasarkan pengumpulan data penelitian dari 1 sumber tersebut, diperoleh data 165 data wilayah bseserta 3 variabel input, selanjutnya data ini digunakan sebagai masukan clustering data penelitian ini selanjutnya di sajikan pada tabel 1 berikut.

Tabel 2. Data Penelitian

o	Nama Wilayah	X1	X2	X3
1	Asahan	38,50	-34,32	33,61
2	Batu bara	4,97	413,93	3,40
3	Binjai	28,06	115,29	18,24
4	Dairi	28,22	2842,94	17,57
5	Deli Serdang	38,05	-87,39	45,38
6	Gunung Sitoli	67,87	-84,25	33,77
7	Humbang Hasundutan	18,59	-98,63	25,77
8	Karo	11,82	-81,06	14,09
9	Labuhan batu	16,18	-87,38	18,41
10	Labuhan Batu Selatan	16,00	-89,40	17,15

Normalisasi

Sebelum diterapkan Algoritma *K-Means* seluruh data dinormalisasikan terlebih dahulu dengan metode Min-Max, agar seluruh variabel berada pada rentang nilai yang sama yaitu antara 0 hingga 1. Normalisasi diperlukan karena ketiga variabel memiliki satuan dan sekala yang

Journal of Science and Social Research ISSN 2615 – 4307 (Print)
Oct 2026, IX (5): 8247 – 8257 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

berbeda beda. Rumus Normalisasi Min-Mix sebagai berikut.

$$X_{norm} = \frac{(X - X_{\min})}{X_{max} - X_{\min}}$$

Berdasarkan Hasil Normalisasi di peroleh hasil nya dari tabel 3 berikut.

Tabel 3. Data Normalisasi

No	Nama Wilayah	X1	X2	X3
1	Asahan	0,438	0,022	0,728
2	Batu Bara	0	0,178	0,030
3	Binjai	0,302	0,073	0,373
4	Dairi	0,304	1	0,357
5	Deli Serdang	0,432	0,004	1
6	Gunung Sitoli	0,822	0,005	0,732
7	Humbang Hasundutan	0,178	0	0,547
8	Karo	0,089	0,006	0,277
9	Labuhan Batu	0,147	0,004	0,377
10	Labuhan Batu Selatan	0,144	0,003	0,348

Nilai hasil normalisasi tersebut kemudian digunakan sebagai masukan proses iterasi *K-Means*.

Penerapan Algoritma K-Means

Setelah data dinormalisasi, algoritma K-Means diterapkan dengan jumlah kluster = 3, yang menampilkan tiga kategori wilayah kasus hipertensi yang tersebar selama 5 tahun terakhir yaitu Tinggi, Sedang dan Rendah.

1 Penentuan Titik Pusat Awal *Centroid*: Pemilihan *centroid* awal pada iterasi pertama dilakukan secara acak dengan mengambil sampel data wilayah yang telah melewati tahap normalisasi *Min-Max*. Pada perhitungan manual ini, dipilih tiga wilayah sebagai representasi titik pusat awal, yaitu:

Tabel 4 Titik Awal *Centroid*

Centroid	**Nama Wilayah**	**X1**	**X2**	**X3**
C1	Asahan	0,438	0,022	0,728
C2	Batu Bara	0	0,178	0,030
C3	Binjai	0,302	0,073	0,373

Pengukuran jarak antara setiap data wilayah dengan titik pusat *centroid* yang masing masih klaster dilakukan menggunakan E*uclidean Distance.* Rumus yang di gunakan pada penelitian ini adalah sebagai berikut.

Dataset 1 Asahan (0.438,0.022,0.728)

$$D = \sqrt{(X_1 - C_1)^2 + (X_2 - C_2)^2 + (X_3 - C_3)^2}$$

$$\sqrt{(0.438 - 0.438)^2 + (0.022 - 0.022)^2 + (0.728 - 0.728)^2} = \sqrt{(0)^2 + (0)^2 + (0)^2 = 0}$$

Jarak dari data (0.438, 0.022, 0.728) ke *cluster* 2 (0, 0.174, 0.030)

$$\sqrt{(0.438 - 0)^2 + (0.022 - 0.174)^2 + (0.728 - 0.030)^2} = \sqrt{(0.438)^2 + (-0.152)^2 + (0.698)^2} = \sqrt{0.1918 + 0.0231 + 0.4872} = \sqrt{0.7021} = 0.838$$

Jarak dari data (0.438, 0.022, 0.728) ke cluster 3 (0.302, 0.073, 0.373):

$$\sqrt{(0.438 - 0.302)^2 + (0.022 - 0.073)^2 + (0.728 - 0.373)^2} = \sqrt{(0.136)^2 + (-0.051)^2 + (0.355)^2} = \sqrt{0.0185 + 0.0026 + 0.1260} = \sqrt{0.1471} = 0.384$$

Hasil dataset Asahan masuk *cluster* 1 karena jarak minimum adalah 0.

Dataset 2 Batu Bara (0, 0.174, 0.030)

$$\sqrt{0 - 0.438)^2 + (0.174 - 0.022)^2 + (0.030 - 0.728)^2} = \sqrt{(-0.438)^2 + (0.152)^2 + (-0.698)^2} = \sqrt{0.1918 + 0.0231 + 0.4872} = \sqrt{0.7021} = 0.838$$

Jarak dari data (0, 0.174, 0.030) ke cluster 2 (0, 0.174, 0.030):

$$\sqrt{(0 - 0)^2 + (0.174 - 0.174)^2 + (0.030 - 0.030)^2} = \sqrt{(0)^2 + (0)^2 + (0)^2 = 0}$$

Jarak dari data (0, 0.174, 0.030) ke cluster 3 (0 .302, 0.073, 0.373):

$$\sqrt{(0 - 0.302)^2 + (0.174 - 0.073)^2 + (0.030 - 0.373)^2} = \sqrt{(-0.302)^2 + (0.101)^2 + (-0.343)^2} = \sqrt{0.0912 + 0.0102 + 0.1176} = \sqrt{0.2190} = 0.468$$

Hasil dataset Batu Bara masuk ke *cluster* 2 karena jarak minimum adalah 0

Dataset 3 Binjai (0.302, 0.073, 0.373)

Journal of Science and Social Research ISSN 2615 – 4307 (Print)
Oct 2026, IX (5): 8247 – 8257 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

Jarak dari data (0.304, 1.000, 0.357) ke *cluster* 2 (0.000, 0.174, 0.030):

$$\sqrt{(0.302-0.438)^2+(0.073-0.022)^2+(0.373-0.728)^2} = \sqrt{(-0.136)^2+(0.051)^2+(-0.355)^2} = \sqrt{0.0185+0.0026+0.1260} = \sqrt{0.1471} = 0.384$$

Jarak dari data (0.302, 0.073, 0.373) ke *cluster* 2 (0, 0.174, 0.030):

$$\sqrt{(0.302-0)^2+(0.073-0.174)^2+(0.373-0.030)^2} = \sqrt{(0.302)^2+(-0.101)^2+(0.343)^2} = \sqrt{0.0912+0.0102+0.1176} = \sqrt{0.2190} = 0.468$$

Jarak dari data (0.302, 0.073, 0.373) ke cluster 3 (0.302, 0.073, 0.373):

$$\sqrt{(0.302-0.302)^2+(0.073-0.073)^2+(0.373-0.373)^2} = \sqrt{(0)^2+(0)^2+(0)^2} = 0$$

Hasil dataset Binjai masuk ke cluster 3 karena jarak minimum adalah 0

Tabel 5. Hasil Iterasi 1

No	Nama Wilayah	X1_n	X2_n	X3_n	d(C1)	d(C2)	d(C3)	Jarak Min	Kluster
1	Asahan	0.438	0.022	0.728	0	0.838	0.384	0	C1
2	Batu Bara	0	0.178	0.030	0.838	0	0.468	0	C2
3	Binjai	0.302	0.073	0.373	0.384	0.468	0	0	C3
4	Dairi	0.304	1	0.357	1.054	0.938	0.927	0.927	C3
5	Deli Serdang	0.432	0.004	1	0.272	1.075	0.644	0.272	C1
6	Gunung Sitoli	0.822	0.005	0.732	0.384	1.094	0.636	0.384	C1
7	Humbang Hasundutan	0.178	0	0.547	0.318	0.574	0.226	0.226	C3
8	Karo	0.089	0.006	0.277	0.571	0.312	0.243	0.243	C3
9	Labuhan Batu	0.147	0.004	0.377	0.458	0.393	0.155	0.155	C3
10	Labuhan Batu Selatan	0.144	0.003	0.348	0.481	0.389	0.175	0.175	C3

Setelah selesai mengelompokkan bedasarkan kluster nya berikut adalah nilai centroid baru digunakan untuk iterasi 2:

Tabel 6. Centroid Baru

1	C1 baru	0.564	0.010	0.820
2	C2 Baru	0.000	0.187	0.030
3	C3 Baru	0.194	0.181	0.380

Iterasi 2 Perhitungan Jarak Euclidean Distance

Dataset 1: Asahan (0.438, 0.022, 0.728)

$$\sqrt{(0.438-0.564)^2+(0.022-0.010)^2+(0.728-0.820)^2} = \sqrt{(-0.126)^2+(0.012)^2+(-0.092)^2} = \sqrt{0.0159+0.0001+0.0085} = \sqrt{0.0245} = 0.157$$

Jarak ke cluster 2 (0.000, 0.178, 0.030):

$$\sqrt{(0.438-0.000)^2+(0.022-0.178)^2+(0.728-0.030)^2} = \sqrt{(0.438)^2+(-0.156)^2+(0.698)^2} = \sqrt{0.1918+0.0243+0.4872} = \sqrt{0.7033} = 0.839$$

Jarak ke cluster 3 (0.194, 0.181, 0.380):

$$\sqrt{(0.438-0.194)^2+(0.022-0.181)^2+(0.728-0.380)^2} = \sqrt{(0.244)^2+(-0.159)^2+(0.348)^2} = \sqrt{0.0595+0.0253+0.1211} = \sqrt{0.2059} = 0.454$$

Hasil: dataset Asahan masuk ke cluster 1 karena jarak minimum adalah 0.157

Dataset 2: Batu Bara (0.000, 0.178, 0.030)

$$\sqrt{(0.000-0.564)^2+(0.178-0.010)^2+(0.030-0.820)^2} = \sqrt{(-0.564)^2+(0.168)^2+(-0.790)^2} = \sqrt{0.3181+0.0282+0.6241} = \sqrt{0.9704} = 0.985$$

Jarak ke cluster 2:

$$\sqrt{(0.000-0.000)^2+(0.178-0.178)^2+(0.030-0.030)^2} = \sqrt{0^2+0^2+0^2} = 0$$

Jarak ke cluster 3:

$$\sqrt{(0.000-0.194)^2+(0.178-0.181)^2+(0.030-0.380)^2} = \sqrt{(-0.194)^2+(-0.003)^2+(-0.350)^2} = \sqrt{0.0376+0.0000+0.1225} = \sqrt{0.1601} = 0.400$$

Hasil: dataset Batu Bara masuk ke cluster 2 karena jarak minimum adalah 0

Dataset 3: Binjai (0.302, 0.073, 0.373)

$$\sqrt{(0.302-0.564)^2+(0.073-0.010)^2+(0.373-0.820)^2} = \sqrt{(-0.262)^2+(0.063)^2+(-0.447)^2} = \sqrt{0.0686+0.0040+0.1998} = \sqrt{0.2724} = 0.522$$

Jarak ke cluster 2:

$$\sqrt{(0.302-0.000)^2+(0.073-0.178)^2+(0.373-0.030)^2} = \sqrt{(0.302)^2+(-0.105)^2+(0.343)^2} = \sqrt{0.0912+0.0110+0.1176} = \sqrt{0.2198} = 0.469$$

Jarak ke cluster 3:

No	Nama Wilayah	Cluster Akhir	Kategori
1	Asahan	C1	Tinggi
2	Batu Bara	C2	Rendah
3	Binjai	C3	Sedang
4	Dairi	C3	Sedang
5	Deli Serdang	C1	Tinggi
6	Gunung Sitoli	C1	Tinggi
7	Humbang Hasundutan	C3	Sedang
8	Karo	C3	Sedang
9	Labuhan Batu	C3	Sedang
10	Labuhan Batu Selatan	C3	Sedang

$$\sqrt{(0.302-0.194)^2+(0.073-0.181)^2+(0.373-0.380)^2} = \sqrt{(0.108)^2+(-0.108)^2+(-0.007)^2} = \sqrt{0.0117+0.0117+0.0000} = \sqrt{0.0234} = 0.153$$

Hasil: dataset Binjai masuk ke cluster 3 karena jarak minimum adalah 0.153

Journal of Science and Social Research ISSN 2615 – 4307 (Print)
Oct 2026, IX (5): 8247 – 8257 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

Tabel 7. Hasil Iterasi 2

No	Nama Wilayah	X1_n	X2_n	X3_n	d(C1)	d(C2)	d(C3)	Jarak min	Kluster
1	Asahan	0.438	0.022	0.728	0.157	0.839	0.454	0.157	C1
2	Batu Bara	0	0.178	0.030	0.985	0	0.400	0	C2
3	Binjai	0.302	0.073	0.373	0.522	0.469	0.153	0.153	C3
4	Dairi	0.304	1	0.357	1.123	0.935	0.827	0.827	C3
5	Deli Serdang	0.432	0.004	1	0.223	1.076	0.687	0.223	C1
6	Gunung Sitoli	0.822	0.005	0.732	0.273	1.095	0.741	0.273	C1
7	Humbang Hasundutan	0.178	0	0.547	0.473	0.575	0.247	0.247	C3
8	Karo	0.089	0.006	0.277	0.721	0.314	0.228	0.228	C3
9	Labuhan Batu	0.147	0.004	0.377	0.608	0.415	0.183	0.183	C3
10	Labuhan Batu Selatan	0.144	0.003	0.348	0.632	0.390	0.188	0.188	C3

Setelah di lakukan perhitungan iterasi 2, pengelompokkan klaster tidak mengalami perubahan di bandingkan sebelumnya. Kondisi ini menunjukkan bahwa algoritma K-Means telah mencapai konvergensi sehingga proses iterasi dihentikan. Berikut adalah hasil dari pengelompokkan wilayah menjadi 3 kluster yang di sajikan pada tabel 8 dibawah.

Tabel 8 Hasil Pengelompokkan

No	Nama Wilayah	Cluster Akhir	Kategori
1	Asahan	C1	Tinggi
2	Batu Bara	C2	Rendah
3	Binjai	C3	Sedang
4	Dairi	C3	Sedang
5	Deli Serdang	C1	Tinggi
6	Gunung Sitoli	C1	Tinggi
7	Humbang Hasundutan	C3	Sedang
8	Karo	C3	Sedang
9	Labuhan Batu	C3	Sedang
10	Labuhan Batu Selatan	C3	Sedang

Evaluasi Hasil dengan Menggunakan Silhouette Score

Evaluasi model dilakukan dengan metode *Silhouette Score* berdasarkan data hasil normalisasi pada tabel 3 metode ini digunakan karna mampu menghitung dua aspek yaitu tingat kohesi dan separasi. Berikut adalah rumus yang digunakan pada metode *Shilhouette Score*

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

Keterangan:
a(i) = rata-rata jarak data ke-i terhadap seluruh anggota cluster yang sama
b(i) = rata-rata jarak minimum data ke-i terhadap anggota cluster lain yang terdekat.
Berikut disajikan perhitungan Silhouette Score secara manual untuk 10 wilayah.

Wilayah 1 Asahan (*Cluster* C1)
a(i) = rata-rata jarak Asahan terhadap sesama anggota *Cluster* C1 = 0.3285
b(i) = nilai minimum dari: rata-rata ke C2 = 0.8386; rata-rata ke C3 = 0.5438
b(i) = 0.5438
max(a(i), b(i)) = max(0.3285, 0.5438) = 0.5438
s(Asahan) = (0.5438 - 0.3285) / 0.5438 = 0.3959

Wilayah 2 Batu Bara (*Cluster* C2)
a(i) = rata-rata jarak Batu Bara terhadap sesama anggota *Cluster* C2 = 0
b(i) = nilai minimum dari: rata-rata ke C1 = 0.9998; rata-rata ke C3 = 0.4431
b(i) = 0.4431
max(a(i), b(i)) = max(0, 0.4431) = 0.4431
s(Batu Bara) = 0 (Ketetapan Silhouette untuk klaster beranggota tunggal)

Wilayah 3 Binjai (*Cluster* C3)
a(i) = rata-rata jarak Binjai terhadap sesama anggota *Cluster* C3 = 0.3480
b(i) = nilai minimum dari: rata-rata ke C1 = 0.4688; rata-rata ke C2 = 0.4727
b(i) = 0.4688
max(a(i), b(i)) = max(0.3480, 0.4688) = 0.4688
s(Binjai) = (0.4688 - 0.3480) / 0.4688 = 0.2576

Tabel 9. Hasil Evaluasi dengan Silhouette Score

No	Nama Wilayah	Klaster	Kohesi a(i)	Separasi b(i)	Nilai Max	*Silhouette Score* s(i)	Keterangan	Kluster
1	Asahan	C1	0.3285	0.5438	0.5438	0.3959	Tepat	Tinggi
2	Batu Bara	C2	0.0000	0.4431	0.4431	0.000	Klaster Tunggal	Rendah
3	Binjai	C3	0.3480	0.4688	0.4688	0.2576	Tepat	Sedang
4	Dairi	C3	0.9982	0.9300	0.9982	-0.0683	Perbatasan	Sedang
5	Deli Serdang	C1	0.3729	0.6015	0.6015	0.3800	Tepat	Tinggi
6	Gunung Sitoli	C1	0.4288	0.6502	0.6502	0.3405	Tepat	Tinggi
7	Humbang Hasundutan	C3	0.3821	0.5600	0.5600	0.3176	Tepat	Sedang
8	Karo	C3	0.3506	0.3110	0.3506	-0.1129	Perbatasan	Sedang
9	Labuhan Batu	C3	0.2992	0.4123	0.4123	0.2743	Tepat	Sedang
10	Labuhan Batu Selatan	C3	0.3011	0.3872	0.3872	0.2223	Tepat	Sedang

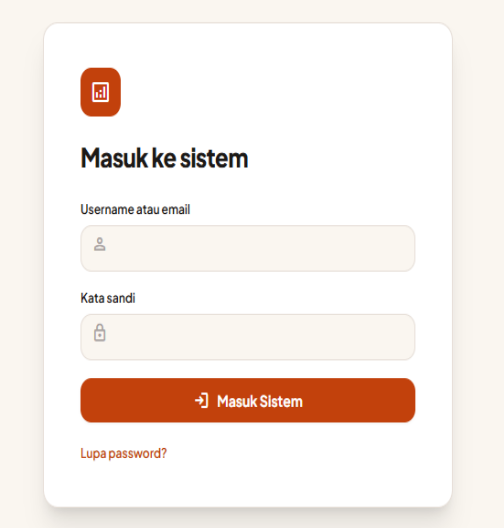
Journal of Science and Social Research ISSN 2615 – 4307 (Print)
Oct 2026, IX (5): 8247 – 8257 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

Implementasi

Sistem pengelompokan wlayah kabuoaten/kota di Sumatera Utara telah berhasil di implementasikan sebagai aplikasi berbasis web menggunakan bahasa pemrograman PHP dan sistem manajemen basis data MySQL. Desain antarmuka menggunakan framework Bootstrap 5 dengan tema warna oren warn yang hangat. Sistem menerapkan algoritma K-Means secara otomatis berdasarkan tiga variabel input yaitu rata-rata kasus, pertumbuhan suatu kasus dan fluktuasi dan sistem menampilkan 6 halaman mulai dari login hingga hasil dan laporan yang akan di tampilakan dibawah.

Tampilan Halaman Login

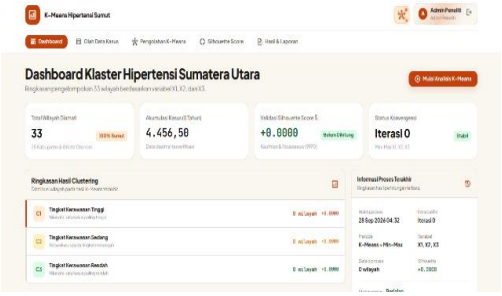
Halaman login antarmuka pertama dalam sistem ini, ketika pengguna mengakses alamat situs yang akan menampilkan card dengan judul masuk ke sistem dengan isian kolom username, password, masuk ke sistem dan lupa password yang akan mengubah password jika lupa. Jika berhasil sistem akan masuk ke halaman Dashboard.



Gambar 4 Tampilan Halaman Login

Tampilan Halaman Dashboard

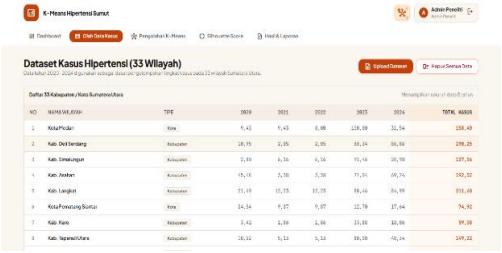
Halaman Dasboard antarmuka kedua dalam sistem ini. Ketika berhasil login halaman dashboard menampilkan 4 card informasi, ringkasan clustering, informasi proses terakhir dan akses cepat.



Gambar 5 Tampilan Halaman Dasboard

Tampilan Halaman Olah Data Kasus

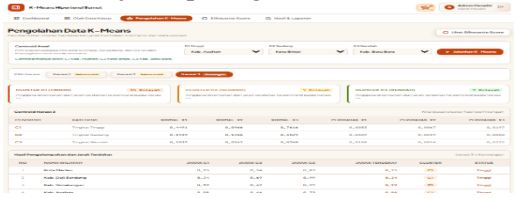
Halaman Olah Data Kasus antarmuka ketiga dalam sistem ini . Ketika pengguna menekan tombol Olah Data Kasus sistem akan menampilakan tombol Uplode data dan Hapus data ketika sudah di uplode sistem akan menampilkan datanya.



Gambar 6 Halaman Olah Data Kasus

Tampilan Halaman Pengolahan K-Means

Halaman Pengolahan K-Means antarmuka keempat dalam sistem ini. Ketika pengguna menekan tombol Pengolahan K-Means, sistem menampilkan card untuk menentukan centroid awal secara manual dan juga bisa secara otomatis, tombol jalankan K-Means, card iterasi, 3 card cluster dan juga tampilan data yang sudah di oleh K-Means dengan jarak Euclidean Distance yang akan di sajikan pada gambar 7 berikut.

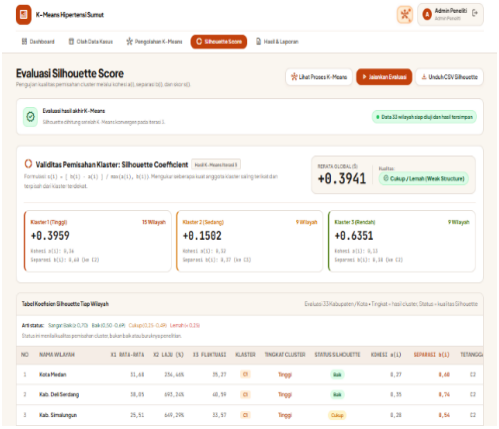


Gambar 7 Tampilan Halaman Pengolahan Data K-Means

Journal of Science and Social Research ISSN 2615 – 4307 (Print)
Oct 2026, IX (5): 8247 – 8257 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

Tampilan Halaman Shilhouette Score

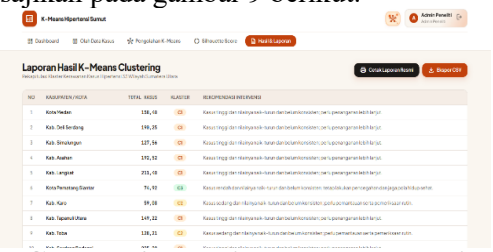
Halaman Shilhouette Score antarmuka kelima dalam sistem ini. Ketika pengguna menekan tombol Shilhouette Score sistem menampilan tombol lihat proses K-Means, jalankan evaluasi dan unduh csv. Jika di tombol K-Means di jalankan pada halaman ini akan langsung otomatis terisi dengan data yang sudah di uplode dengan menampilan card evaluasi hasil akhir, 3 card kalster C1, C2, C3, rerata global dan juga tabel hasil dari evaluasi Shilhouette score**.**



Gambar 8 Tampilan Halaman Silhouette Score

Tampilan Halaman Hasil dan Laporan

Halaman Hasil dan Laporan antarmuka keenam dalam sistem ini. Ketika pengguna menekan tombol hasil dan laporan sistem menampilan tombol cetak laporan untuk membuka jendela ketika ingin mengunduh dengan format pdf dan di sebelah nya tombol eksport csv jika ingin mengunduh format csv dan di bawah nya menampilkan tabel data yang sudah di olah dengan K-Means dan di evaluasi dengan shilhouette score yang di sajikan pada gambar 9 berikut.



Gambar 9 Tampilan Halaman Hasil dan Laporan

SIMPULAN

Berdasarkan tahapan penelitian, proses perhitungan manual, dan perancangan sistem pengelompokan menggunakan algoritma K-Means yang telah diuraikan pada bab-bab sebelumnya, dapat ditarik beberapa kesimpulan utama sebagai berikut:

1. Keberhasilan Implementasi K-Means: Algoritma K-Means dengan penentuan nilai k=3 telah berhasil diterapkan untuk memetakan data kasus wilayah ke dalam tiga karakteristik klaster yang jelas, yaitu klaster Tingkat Tinggi (C1), Tingkat Rendah (C2), dan Tingkat Sedang (C3).
2. Kestabilan Proses Iterasi: Melalui perhitungan manual menggunakan sampel 10 wilayah dan tahap pra-pemrosesan normalisasi Min-Max, proses K-Means terbukti efisien dan langsung mencapai titik konvergen (kestabilan anggota klaster) pada Iterasi ke-2.
3. Validasi Kinerja Model yang Positif: Hasil pengujian kualitas klaster menggunakan metode Silhouette Score memberikan nilai rata-rata keseluruhan (Global Silhouette Index) sebesar 0.2007. Karena nilai tersebut berada pada rentang positif, secara matematis terbukti bahwa struktur klaster yang dihasilkan sudah wajar, valid, dan penempatan wilayah ke dalam masing-masing klasternya sudah cukup tepat.
4. Pembangunan Sistem Terintegrasi: Logika perhitungan algoritma berhasil diimplementasikan ke dalam sistem berbasis web menggunakan PHP Native, yang mampu memproses dan mengelompokkan skala data yang lebih besar (33 Kabupaten/Kota di Sumatera Utara) secara otomatis

DAFTAR PUSTAKA

Fadillah, A. (2024). *Pengembangan Aplikasi Mobile Berbasis Flutter untuk Pengenalan Objek dalam Video Real-Time*. 1–11.

Faqih, A., & Basysyar, F. M. (2022). *Sistem Informasi Kepegawaian Mengunakan Metode Agile Development di CV . Angkasa Raya*. *12*, 30–45. Https://doi.org/10.34010/jati.v12i1

Fithryani, N. M., Dikananda, A. R., & Rohman, D. (2025). *ALGORITMA K-MEANS UNTUK*. *13*(1), 997–1003.

Hasanah, N. A., Rodianto, R., Yuliadi, Y., & Sumbawa, U. T. (2025). *Rancang Bangun Aplikasi Sistem Pakar Menggunakan Metode Forward Chaining untuk Mendiagnosa Penyakit pada Ayam 1,2*. *12*(2), 425–441.

Ina, W. T., Mesakh, Y., & Pella, S. E. (2022). *Klusterisasi penyakit endemis pada kecamatan sabu barat , kabupaten sabu raijua menggunakan algoritma k-means*. *Xi*(1), 39–44. Https://doi.org/10.35508/jme.v0i0.6508

Laelatul, & Lorenza. (2025). *Jurnal Processor Penerapan Clustering*. *20*(1), 44–57.

Maulia, S., Ginting, B. S., & Sihombing, A. (2021). *Implementasi data mining pengelompokan jenis penyakit pasien menggunakan metode clustering (studi kasus : puskesmas sambirejo)*. *5*(1), 71–80.

Meida, T., Tyas, M., & Purnamasari, A. I. (2023). *Penerapan Algoritma K - means dalam Mengelompokkan Demam Berdarah Dengue Berdasarkan Kabupaten*.

Qirom, D. S., Faqih, A., & Dwilestari, G. (2024). *Implementasi algoritma k-means untuk klasterisasi pasien*. *8*(2), 2056–2063.

Sari, indsari, I. P., Jannah, A., Meuraxa, A. M., Syahfitri, A., & Omar, R. (2022). P. S. I. P. D. M. B. W. A. P., Jannah, A., Meuraxa, A. M., Syahfitri, A., & Omar, R. (2022). *Perancangan Sistem Informasi Penginputan Database Mahasiswa Berbasis Web*.

Widiyanto, D. (2022). *Perancangan sistem informasi manajemen inventori berbasis web (studi kasus : smk ypt purworejo)*. *10*(1), 24–31