
OPTIMASI ELBOW PADA METODE K-MENAS DALAM CLUSTERING DATA DARAH PASIEN UNTUK IDENTIFIKASI RESIKO PENYAKIT KRONIS

Pani Irawan¹, Dicky Apdillah²

Universitas Asahan, Kisaran

Email: ¹irawanpani117@gmail.com, ²dickyapdi1404@gmail.com

Abstract: *Chronic diseases such as diabetes, hypertension, and heart disease are major causes of morbidity and mortality in various countries, including Indonesia. Early detection of chronic disease risk is a crucial step in more effective prevention and treatment. One important source of information for identifying these risks is laboratory data, particularly patient blood data, which includes parameters such as glucose levels, cholesterol, blood pressure, hemoglobin, and other indicators. The objective of this study was to design and develop a patient blood data clustering application to identify chronic disease risk using the K-Menas method and the Elbow method using a web programming language. The data used in this study consisted of a sample of 140 individuals. Of the total sample, 140 patients were successfully grouped into four optimal clusters ($K = 4$). This data partitioning process resulted in a varied population distribution within each cluster, reflecting the heterogeneity of clinical and demographic characteristics among the study subjects. Based on the analysis of 140 patients, the population was dominated by the Critical/Chronic group (Cluster 4) at 45.7% (64 patients), followed by the Healthy/Normal Group (Cluster 1) at 30% (42 patients), and the Low/Early Risk (Cluster 2) and High Risk (Cluster 3) groups, each with an equal minority proportion of 12.1% (17 patients).*

Keywords: *Elbow Optimization, K-Menas Method, Clustering, Patient Blood Data, Chronic Disease Risk Identification*

Abstrak: Penyakit kronis seperti diabetes, hipertensi, dan penyakit jantung menjadi penyebab utama morbiditas dan mortalitas di berbagai negara, termasuk Indonesia. Deteksi dini terhadap risiko penyakit kronis merupakan langkah penting dalam upaya pencegahan dan penanganan yang lebih efektif. Salah satu sumber informasi penting untuk mengidentifikasi risiko tersebut adalah data laboratorium, khususnya data darah pasien, yang mencakup parameter-parameter seperti kadar glukosa, kolesterol, tekanan darah, hemoglobin, serta indikator lainnya. Tujuan dalam penelitian ini ialah Untuk merancang dan membuat aplikasi clustering data darah pasien untuk identifikasi risiko penyakit kronis menggunakan Metode K-Menas dan metode elbow dengan bahasa pemrograman web. Data yang digunakan dalam penelitian ini berupa sample sebanyak 140 orang. Dari total sampel penelitian sebanyak 140 pasien berhasil dikelompokkan ke dalam 4 klaster optimal ($K = 4$). Proses partisi data ini menghasilkan sebaran populasi yang bervariasi di setiap klasternya, mencerminkan adanya heterogenitas karakteristik klinis atau demografis di antara subjek penelitian. Berdasarkan hasil analisis terhadap 140 pasien, populasi didominasi oleh kelompok Kritis/Kronis (Klaster 4) sebesar 45,7% (64 pasien), diikuti kelompok Kelompok Sehat/Normal (Klaster 1) sebesar 30% (42 pasien), serta kelompok Risiko Rendah/Awal (Klaster 2) dan Risiko Tinggi (Klaster 3) yang memiliki proporsi minoritas setara masing-masing sebesar 12,1% (17 pasien).

Kata Kunci: Optimasi Elbow, Metode K-Menas, Clustering, Data Darah Pasien, Identifikasi Resiko Penyakit Kronis

PENDAHULUAN

Perkembangan teknologi di bidang kesehatan merupakan salah satu pilar utama dalam transformasi sistem pelayanan medis modern. Seiring dengan kemajuan teknologi informasi, komputasi dan bioteknologi, sektor kesehatan mengalami disrupsi signifikan dalam hal diagnosis, pengobatan, pemantauan, serta manajemen data pasien. Inovasi ini tidak hanya meningkatkan efisiensi pelayanan, tetapi juga mendorong personalisasi dan presisi dalam perawatan kesehatan.

Penyakit kronis seperti diabetes, hipertensi, dan penyakit jantung menjadi penyebab utama morbiditas dan mortalitas di berbagai negara, termasuk Indonesia. Deteksi dini terhadap risiko penyakit kronis merupakan langkah penting dalam upaya pencegahan dan penanganan yang lebih efektif. Salah satu sumber informasi penting untuk mengidentifikasi risiko tersebut adalah data laboratorium, khususnya data darah pasien, yang mencakup parameter-parameter seperti kadar glukosa, kolesterol, tekanan darah, hemoglobin, serta indikator lainnya.

RS. Sapta Medika Indrapura yang beralamatkan di jl. Datuk Umar Pelangki, Indrapura, Tanah Merah Kabupaten Batu Bara sebagai salah satu fasilitas pelayanan kesehatan memiliki jumlah data pemeriksaan darah pasien yang terus meningkat setiap tahunnya. Namun, hingga saat ini pemanfaatan data tersebut masih terbatas pada kebutuhan diagnosis individual pasien dan belum digunakan secara maksimal untuk analisis pola kesehatan pasien secara menyeluruh. Akibatnya, potensi data dalam mengidentifikasi kelompok pasien dengan risiko penyakit kronis tertentu belum dapat dimanfaatkan secara optimal.

Penelitian terdahulu yang pernah dilakukan oleh (Ishak, 2024) dengan judul Optimasi K-Means pada Clustering Penyakit Ibu Hamil Menggunakan Random Forest. Hasil pengujian

menunjukkan bahwa penggunaan tiga atribut ini meningkatkan *Silhouette Score* sebesar 0,21 (dari 0,23 menjadi 0,44), yang mengindikasikan pemisahan *cluster* yang lebih baik, serta menurunkan *Davies-Bouldin Index* sebesar 0,69 (dari 1,50 menjadi 0,81), menunjukkan *cluster* yang lebih kompak dan terpisah dengan baik. Visualisasi *clustering* menggunakan *Principal Component Analysis* (PCA) mendukung hasil ini. Selain itu, perhitungan metode Elbow menunjukkan jumlah *cluster optimal* pada $k=3$, memperkuat kesimpulan bahwa pemilihan *atribut* dan jumlah *cluster* yang tepat meningkatkan kualitas *clustering*. Secara keseluruhan, penelitian ini membuktikan bahwa seleksi fitur berbasis Random Forest mampu mengoptimalkan metode K-Means dalam clustering penyakit pada ibu hamil, yang diharapkan dapat meningkatkan efektivitas diagnosis dan perencanaan perawatan.

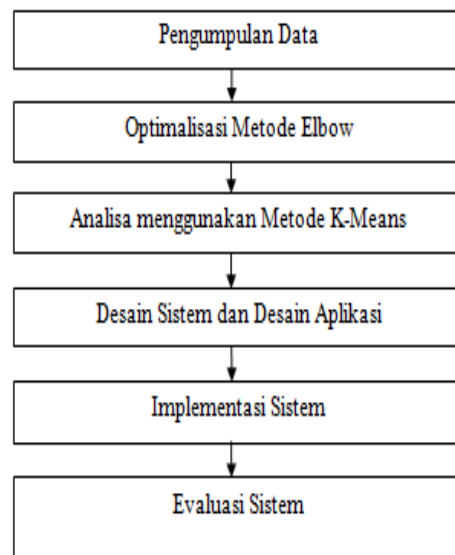
Penelitian terdahulu yang pernah dilakukan oleh (Qusyairi et al., 2024) dengan judul Penerapan K-Means Clustering Dalam Pengelompokan Prestasi Siswa Dengan Optimasi Metode Elbow. Hasil penelitian ini berupa sistem yang dapat mempermudah sekolah dengan cepat dan tepat dalam mengelompokkan prestasi siswa menggunakan k-means yang dioptimasi dengan penggunaan metode elbow dengan perolehan hasil yang didapat yaitu cluster 1 “baik” sebanyak 9 siswa, cluster 2 “sangat baik” sebanyak 6 siswa, cluster 3 “cukup baik” sebanyak 8 siswa dan nilai *davies bouldin* 0,711 dan *avg within centroid distance* 18,821 sehingga bias bisa membantu pihak sekolah dalam menentukan pengelompokan prestasi siswa.

Berdasarkan masalah diatas, diperlukan optimalisasi metode Elbow untuk menentukan jumlah cluster yang lebih akurat dan objektif pada Metode K-Means. Dengan optimalisasi ini diharapkan hasil clustering data darah pasien menjadi lebih representatif, sehingga pengelompokan risiko penyakit

kronis dapat dilakukan dengan lebih tepat. Penerapan pendekatan ini diharapkan mampu mendukung tenaga medis atau pihak rumah sakit dalam melakukan deteksi dini, pengambilan keputusan, dan perencanaan intervensi kesehatan yang lebih efektif.

METODE

Penelitian ini menggunakan pendekatan kuantitatif, di mana proses analisis dilakukan dengan menerapkan metode K-Means yang dioptimalkan menggunakan metode Elbow untuk mengidentifikasi pola risiko penyakit kronis berdasarkan data darah pasien. Pendekatan ini dipilih karena mampu memberikan hasil analisis yang terukur dan objektif melalui pengolahan data numerik hasil pemeriksaan laboratorium. Tahapan penelitian diawali dengan pengumpulan data darah pasien yang diperoleh dari hasil pemeriksaan laboratorium di instansi kesehatan yang menjadi objek penelitian. Data tersebut meliputi beberapa parameter penting seperti kadar glukosa darah, kolesterol total, tekanan darah, hemoglobin, dan asam urat. Seluruh data yang digunakan bersifat sekunder dan telah dianonimkan untuk menjaga kerahasiaan identitas pasien. Dalam penelitian ini proses optimalisasi metode Elbow dilakukan untuk meningkatkan ketepatan dalam menentukan jumlah cluster terbaik pada algoritma K-Means yang digunakan untuk mengelompokkan data darah pasien. Metode Elbow pada dasarnya bekerja dengan menghitung nilai *Within-Cluster Sum of Squares* (WCSS) untuk berbagai jumlah cluster (k). Nilai WCSS akan terus menurun seiring bertambahnya jumlah cluster namun penurunan tersebut akan melambat setelah mencapai titik optimal. Titik perubahan yang tampak seperti “siku” atau “elbow” pada grafik menjadi indikator jumlah cluster yang paling sesuai.



Gambar 1 Kerangka Kerja Penelitian

HASIL DAN PEMBAHASAN

Hasil Penelitian

Analisa Data

Analisa data yang dilakukan peneliti untuk menyelesaikan penelitian ini menggunakan metode kualitatif dengan rancangan studi kasus, alasan penelitian kualitatif digunakan peneliti karena dalam penelitian kualitatif cenderung menggunakan analisis dalam menyelesaikan suatu masalah. Dalam pengumpulan data ini, metode yang penulis gunakan berupa metode observasi yang nantinya digunakan untuk memperoleh data-data yang dibutuhkan dalam pembuatan aplikasi. Adapun data kriteria yang digunakan dalam penelitian ini seperti dibawah ini.

Menentukan data Kriteria

Data uji merupakan faktor pendukung untuk melakukan proses pengujian, data uji yang digunakan dalam penelitian ini sebanyak 6 data kriteria yang diperoleh, adapun data yang digunakan sebagai berikut :

Tabel 1 Data Kriteria

No	Kriteria	Data	Keterangan
1	C1	Chol	Cholesterol Total
2	C2	TG	Trigliserida
3	C3	HDL	High-Density Lipoprotein
4	C4	LDL	Low-Density Lipoprotein
5	C5	Cr	Kreatinin
6	C6	BUN	Blood Urea Nitrogen

Klaster

Dalam menentukan keputusan dari hasil clustering peneliti memiliki 4 klaster seperti:

Tabel 2 Clustering

No	Kelas	Keterangan
1	Kluster 1	Kelompok Sehat (Normal)
2	Kluster 2	Risiko Rendah/Awal
3	Kluster 3	Risiko Tinggi
4	Kluster 4	Kritis/Kronis

Data Training

Setelah data kriteria dan data kluster sudah di dapat langkah selanjutnya menentukan data training. Dalam khusus ini prolehan data training dapat dilihat pada tabel dibawah ini:

Tabel 3 Data Training

NO	Nama	Chol	TG	HDL	LDL	Cr	BUN
1	Irfan Wajdi	4.2	0.9	2.4	1.4	46	4.7
2	Yuni Sulaeman	3.7	1.4	1.1	2.1	62	4.5
3	Emma Adi	4.9	1	0.8	2	46	7.1
4	Tia Corbuzier	2.9	1	1	1.5	24	2.3
5	Olivia Ramadhan	3.6	1.3	0.9	2.1	50	2
6	Farid Mahendra	2.9	0.8	0.9	1.6	47	4.7
7	Bambang Mulyadi	3.8	0.9	2.4	3.7	67	2.6
8	Sulis Wicaksono	3.8	2	2.4	3.8	28	3.6
9	Amanda Widjaja	3.6	0.7	1.7	1.6	55	4.4
10	Fauzan Segara	4	1.1	0.9	2.7	53	3.3
11	Laras Setiawan	4.9	1.3	1.2	3.2	42	3
12	Feni Harahap	4.2	1.7	1.2	2.2	54	4.6
13	Prasetyo Mubarak	4	1.5	1.2	2.2	39	3.5
14	Suci Carlina	3.6	1.1	1	2.1	74	5.5
15	Fikri Ratnasari	5.3	0.8	1.1	4.1	53	5.9
16	Joko Tanjung	5	1.3	1.2	3.3	28	2.2
17	Ramadhan Kurniawan	4.4	0.9	1	1.3	55	3.8
18	Ivana Setiawan	0.5	1.9	1.3	2.8	175	13.5
19	Dedi Munandar	6.2	1	1.1	4.6	73	5
20	Permata Pratama	4.2	1.5	1.2	2.3	62	5.3
....
140	Wulan Hadinata	4.9	1.6	1	3.2	55	3.4

Pengujian

Mengolah Data Dengan Metode elbow dan *Algoritma K-Means*. Setelah menentukan variabel dan data yang ada, selanjutnya akan dilakukan pengolahan terhadap data yang diperoleh dari

pengamatan. Langkah-langkah yang dilakukan adalah dengan membuat normalisasi data. Berikut normalisasi terhadap dataset:

1. Normalisasi Dataset

mencari nilai normalisasi data peneliti menggunakan rumus:

$$x^i = \frac{x_i}{|\max(x)|}$$

Contoh perhitungan untuk menormalisasikan nilai X1 pada penilaian Chol, TG, HDL, LDL, Cr dan BUN.

$$0001 (\text{Chol}) = \frac{4,2}{9,5} = 0,442105263$$

$$0001 (\text{TG}) = \frac{0,9}{5,9} = 0,152542373$$

$$0001 (\text{HDL}) = \frac{2,4}{4} = 0,6$$

$$0001 (\text{LDL}) = \frac{1,4}{4,9} = 0,285714286$$

$$0001 (\text{Cr}) = \frac{46}{344} = 0,13372093$$

$$0001 (\text{BUN}) = \frac{4,7}{22} = 0,213636364$$

Contoh perhitungan untuk menormalisasikan nilai X2 pada penilaian Chol, TG, HDL, LDL, Cr dan BUN.

$$0001 (\text{Chol}) = \frac{3,7}{9,5} = 0,389473684$$

$$0001 (\text{TG}) = \frac{1,4}{5,9} = 0,237288136$$

$$0001 (\text{HDL}) = \frac{1,1}{4} = 0,275$$

$$0001 (\text{LDL}) = \frac{2,1}{4,9} = 0,428571429$$

$$0001 (\text{Cr}) = \frac{62}{344} = 0,180232558$$

$$0001 (\text{BUN}) = \frac{4,5}{22} = 0,204545455$$

Perhitungan Algoritma Elbow

Dalam algoritma Elbow, rumus utama yang digunakan untuk menentukan jumlah cluster (K) yang optimal adalah WCSS (*Within-Cluster Sum of Squares*). WCSS mengukur variasi (jarak) data di dalam sebuah cluster. Semakin kecil nilai WCSS, semakin rapat data di dalam cluster tersebut.

Rumus yang digunakan adalah Sebagai Berikut:

$$\text{Cluster} - 2 \quad (0.44210526315789 - 0.45468331846566)^2 + (0.056603773584906 - 0.15813879117365)^2 + (0.542857142814 - 0.20121065)^2 + (0.239130 - 0.48784082535004)^2 + (0.11834319526627 - 0.1590362049)^2 + (0.135 - 0.11292372881356)^2 = 0.19119003195568$$

$$\text{Cluster} - 2 \quad (0.38947368 - 0.45468331)^2 + (0.1509433962 - 0.1581387)^2 + (0.17142857 - 0.20121065375)^2 + (0.39130434782609 - 0.487840)^2 + (0.16568047337278 - 0.15903620499448)^2 + (0.125 - 0.11292372)^2 = 0.014700316656498$$

$$\text{Cluster} - 2 \quad (0.51578947368421 - 0.45468331846566)^2 + (0.075471698113208 - 0.15813879117365)^2 + (0.085714285714286 - 0.20121065375303)^2 + (0.3695652173913 - 0.48784082535004)^2 + (0.11834319526627 - 0.15903620)^2 + (0.255 - 0.11292372881356)^2 = 0.059737928823789$$

$$\text{Cluster} - 2 \quad (0.30526315789474 - 0.45468331846566)^2 + (0.075471698113208 - 0.15813879117365)^2 + (0.14285714285714 - 0.20121065375303)^2 + (0.26086956521739 - 0.48784082535004)^2 + (0.053254437869822 - 0.15903620499448)^2 + (0.015 - 0.11292372881356)^2 = 0.104860156$$

$$\text{Cluster} - 2 \quad (0.37894736842105 - 0.45468331846566)^2 + (0.13207547169811 - 0.15813879117365)^2 + (0.11428571428571 - 0.20121065375303)^2 + (0.39130434782609 - 0.48784082535004)^2 = 0.104860156$$

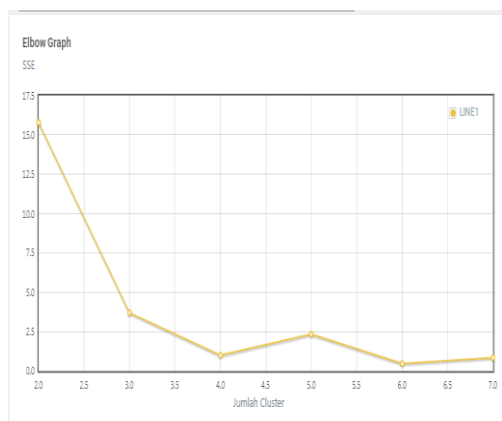
$$\begin{aligned} & 0.48784082535004)^2+(0.13 \\ & 01775147929 \quad - \\ & 0.15903620499448)^2+(0 \quad - \\ & 0.11292372881356)^2=0.03 \\ & 687505987468 \\ \text{Cluster} \quad - \quad & 2 \quad (0.30526315789474 \quad - \\ & 0.45468331846566)^2+(0.03 \\ & 7735849056604 \quad - \\ & 0.15813879117365)^2+(0.11 \\ & 428571428571 \quad - \\ & 0.20121065375303)^2+ \\ & (0.28260869565217 \quad - \\ & 0.48784082535004)^2+(0.12 \\ & 130177514793 \quad - \\ & 0.15903620499448)^2+(0.13 \\ & 5 \quad - \quad 0.11292372881356)^2 \\ & =0.088410673 \\ \text{Cluster} \quad - \quad & 2 \quad (0.4 \quad - \\ & 0.45468331846566)^2+(0.05 \\ & 6603773584906 \quad - \\ & 0.15813879 \quad)^2+ \\ & (0.54285714285714 \quad - \end{aligned}$$

$$\begin{aligned} & 0.20121065375303)^2+(0.73 \\ & 913043478261 \quad - \\ & 0.48784082535004)^2+(0.18 \\ & 047337278107 \quad - \\ & 0.15903620499448)^2+ \\ & (0.03 \quad - \\ & 0.11292372881356)^2=0.20 \\ & 050431340416 \end{aligned}$$

Dalam perhitungan yang menggunakan algoritma elbow, peneliti hanya menampilkan perhitungan cluster 2 sampai cluster 3. Dan dari perhitungan cluster 2 sampai ke 7 terlihat bahwa jumlah cluster yang terbaik adalah cluster 4 sehingga dalam penelitian ini peneliti menggunakan 4 cluster dalam mendeteksi Resiko Penyakit Kronis. Berikut ini tabel perbandingan dari tiap-tiap cluster.

Tabel 6 Hasil Perhitungan Elbow

No	Cluster	Nilai	Selisi
1	Cluster 2	15.785996178616	15.785996178616
2	Cluster 3	12.086843306568	3.6991528720479
3	Cluster 4	11.082334938334	1.0045083682342
4	Cluster 5	8.7427318058036	2.3396031325302
5	Cluster 6	8.2640334629191	0.4786983428845
6	Cluster 7	7.4119191612012	0.85211430171798



Gambar 2 Hasil Algoritma Elbow

1. Nilai k didapat dari hasil perhitungan algoritma elbow sebagai jumlah cluster yang terbentuk.
 - a. Cluster 1 = Kelompok Sehat (Normal)
 - b. Cluster 2 = Risiko Rendah/Awal
 - c. Cluster 3 = Risiko Tinggi
 - d. Cluster 4 = Kritis/Kronis
2. Membangkitkan nilai random untuk pusat cluster awal (*centroid*)

Perhitungan Algoritma K-Means

Tabel 7 Nilai Centroid Awal

No	Kelas	C1	C2	C3	C4	C5	C6
1	Kluster 1	0.432781	0.126684	0.196734	0.522360	0.154015	0.11757
		95488722	6361186	69387755	2484472	21555368	1428571
2	Kluster 2	0.472180	0.193530	0.199183	0.481987	0.128148	0.09685

No	Kelas	C1	C2	C3	C4	C5	C6
		45112782	99730458	67346939	5776397 5	77430262	7142857 143
3	Kluster 3	0.435187 96992481	0.338544 47439353	0.191836 73469388	0.495031 0559006 2	0.228317 83601014	0.18828 5714285 71
4	Kluster 3	0.459548 87218045	0.242587 60107817	0.189387 75510204	0.459627 3291925 5	0.189940 82840237	0.136

Menghitung jarak setiap data input terhadap masing-masing *centroid* menggunakan rumus jarak Euclidean (*Euclidean Distance*) hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*. Berikut adalah persamaan *Euclidian Distance*:

$$d(x_i, u_j) = \sqrt{\sum (x_i - u_j)^2}$$

Irfan Wajdi

$$\begin{aligned} & \sqrt{(0.4421 - 0.4328)^2 + (0.0566 - 0.1267)^2 + (0.5429 - 0.1967)^2 + (0.2391 - 0.5224)^2 + (0.1183 - 0.154)^2 + (0.135 - 0.1176)^2} \\ &= \sqrt{0.0001 + 0.0049 + 0.1198 + 0.0802 + 0.0013 + 0.0003} \\ &= \sqrt{0.2066} \\ &= 0.4545 \\ & \sqrt{(0.4421 - 0.4722)^2 + (0.0566 - 0.1935)^2 + (0.5429 - 0.1992)^2 + (0.2391 - 0.482)^2 + (0.1183 - 0.1281)^2 + (0.135 - 0.0969)^2} \\ &= \sqrt{0.0009 + 0.0187 + 0.1181 + 0.0590 + 0.0001 + 0.0015} \\ &= \sqrt{0.1983} \\ &= 0.4453 \\ & \sqrt{(0.4421 - 0.4352)^2 + (0.0566 - 0.3385)^2 + (0.5429 - 0.1918)^2 + (0.2391 - 0.495)^2 + (0.1183 - 0.2283)^2 + (0.135 - 0.1883)^2} \\ &= \sqrt{0.0000 + 0.0795 + 0.1232 + 0.0655 + 0.0121 + 0.0028} \\ &= \sqrt{0.2832} \\ &= 0.5321 \\ & \sqrt{(0.4421 - 0.4595)^2 + (0.0566 - 0.2426)^2 + (0.5429 - 0.1894)^2 + (0.2391 - 0.4596)^2 + (0.1183 - 0.1899)^2 + (0.135 - 0.136)^2} \\ &= \sqrt{0.0003 + 0.0346 + 0.1249 + 0.0486 + 0.0051 + 0.0000} \\ &= \sqrt{0.2136} \end{aligned}$$

$$= 0.4621$$

Yuni Sulaeman

$$\begin{aligned} & \sqrt{(0.3895 - 0.4328)^2 + (0.1509 - 0.1267)^2 + (0.1714 - 0.1967)^2 + (0.3913 - 0.5224)^2 + (0.1657 - 0.154)^2 + (0.125 - 0.1176)^2} \\ &= \sqrt{0.0019 + 0.0006 + 0.0006 + 0.0172 + 0.0001 + 0.0001} \\ &= \sqrt{0.0205} \\ &= 0.1431 \\ & \sqrt{(0.3895 - 0.4722)^2 + (0.1509 - 0.1935)^2 + (0.1714 - 0.1992)^2 + (0.3913 - 0.482)^2 + (0.1657 - 0.1281)^2 + (0.125 - 0.0969)^2} \\ &= \sqrt{0.0068 + 0.0018 + 0.0008 + 0.0082 + 0.0014 + 0.0008} \\ &= \sqrt{0.0198} \\ &= 0.1409 \\ & \sqrt{(0.3895 - 0.4352)^2 + (0.1509 - 0.3385)^2 + (0.1714 - 0.1918)^2 + (0.3913 - 0.495)^2 + (0.1657 - 0.2283)^2 + (0.125 - 0.1883)^2} \\ &= \sqrt{0.0021 + 0.0352 + 0.0004 + 0.0108 + 0.0039 + 0.0040} \\ &= \sqrt{0.0564} \\ &= 0.2375 \\ & \sqrt{(0.3895 - 0.4595)^2 + (0.1509 - 0.2426)^2 + (0.1714 - 0.1894)^2 + (0.3913 - 0.4596)^2 + (0.1657 - 0.1899)^2 + (0.125 - 0.136)^2} \\ &= \sqrt{0.0049 + 0.0084 + 0.0003 + 0.0047 + 0.0006 + 0.0001} \\ &= \sqrt{0.0190} \\ &= 0.1379 \end{aligned}$$

Setelah proses perhitungan diatas selesai peneliti mengelompokan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil). Data dibawah ini merupakan data keseluruhan.

Tabel 8 Centroid (Jarak Terkecil)

No	Nama	Cluster
1	Irfan Wajdi	2
2	Yuni Sulaeman	4
3	Emma Adi	1
4	Tia Corbuzier	2
5	Olivia Ramadhan	2
6	Farid Mahendra	1
7	Bambang Mulyadi	1
8	Sulis Wicaksono	1
9	Amanda Widjaja	1
10	Fauzan Segara	1
11	Laras Setiawan	1
12	Feni Harahap	4
...
140	Wulan Hadinata	1

rata *cluster* yang bersangkutan dengan menggunakan rumus:

$$u_j^{(t+1)} = \frac{1}{N_{sj}} \sum_{j \in J} x_j$$

Kluster 1 (Cholesterol Total) = 0.46748120300752

Kluster 1 (Trigliserida) = 0.10882749326146

Kluster 1 (High-Density Lipoprotein) = 0.19030612244898

Kluster 1 (Low-Density Lipoprotein) = 0.60170807453416

Kluster 1 (Kreatinin) = 0.15046491969569

Kluster 1 (Blood Urea Nitrogen) = 0.11357142857143

Lakukan perhitungan terhadap kluster 2 dan kluster 4. Sehingga hasilnya dapat dilihat pada tabel dibawah ini.

Memperbaharui nilai Nilai centroid baru di peroleh dari rata-

Tabel 9 Centroid Literasi 2

No	Kelas	C1	C2	C3	C4	C5	C6
1	Kluster 1	0.467481 20300752	0.108827 49326146	0.190306 12244898	0.601708 07453416	0.150464 91969569	0.1135714 2857143
2	Kluster 2	0.481203 0075188	0.153638 81401617	0.259863 94557823	0.417184 26501035	0.106649 76049591	0.0671428 57142857
3	Kluster 3	0.436090 22556391	0.520215 63342318	0.175510 20408163	0.514751 55279503	0.264687 23584108	0.2435714 2857143
4	Kluster 3	0.414135 33834586	0.218867 9245283	0.176326 53061224	0.334161 49068323	0.183939 13778529	0.1218571 4285714

Melakukan perulangan dari langkah 3 hingga 5, sampai anggota tiap cluster tidak ada yang berubah. Setelah proses perhitungan literasi ke 2 diatas selesai peneliti mengelompokan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil).

Irfan Wajdi

$$\begin{aligned} &\sqrt{(0.4421 - 0.4675)^2 + (0.0566 - 0.1088)^2 + (0.5429 - 0.1903)^2 + (0.2391 - 0.6017)^2 + (0.1183 - 0.1505)^2 + (0.135 - 0.1136)^2} \\ &= \sqrt{(0.0006 + 0.0027 + 0.1243 + 0.1315 + 0.0010 + 0.0005)} \\ &= \sqrt{(0.2606)} \\ &= 0.5105 \\ &\sqrt{(0.4421 - 0.4812)^2 + (0.0566 - 0.1536)^2 + (0.5429 - 0.2599)^2 + (0.2391 - 0.4172)^2 + (0.1183 - 0.1066)^2 + (0.135 - 0.0671)^2} \\ &= \sqrt{(0.0015 + 0.0094 + 0.0801 + 0.0317 + 0.0001 + 0.0046)} \\ &= \sqrt{(0.1275)} \\ &= 0.3570 \\ &\sqrt{(0.4421 - 0.4361)^2 + (0.0566 - 0.5202)^2 + (0.5429 - 0.1755)^2 + (0.2391 - 0.5148)^2 + (0.1183 - 0.2647)^2 + (0.135 - 0.2436)^2} \\ &= \sqrt{(0.0000 + 0.2149 + 0.1349 + 0.0760 + 0.0214 + 0.0118)} \\ &= \sqrt{(0.4591)} \\ &= 0.6776 \\ &\sqrt{(0.4421 - 0.4141)^2 + (0.0566 - 0.2189)^2 + (0.5429 - 0.1763)^2 + (0.2391 - 0.3342)^2 + (0.1183 - 0.1839)^2 + (0.135 - 0.1219)^2} \\ &= \sqrt{(0.0008 + 0.0263 + 0.1343 + 0.0090 + 0.0043 + 0.0002)} \end{aligned}$$

$$= \sqrt{(0.1750)}$$

$$= 0.4183$$

Memperbaharui nilai Nilai centroid baru di peroleh dari rata-rata *cluster* yang bersangkutan dengan menggunakan rumus:

$$u_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in i} x_j$$

Tabel dibawah ini merupakan nilai centroid baru pada literasi ke 9, peneliti langsung menuju literasi ke 9 dikarenakan proses yang panjang jika peneliti membuat seluruh perubahan literasi. Cara mendapatkan nilai centroid baru pada literasi ke 9 langkahnya sama dengan literasi ke 2 sehingga peneliti menampilkan pada tabel dibawah ini.

Tabel 10 Centroid Literasi 9

No	Kelas	C1	C2	C3	C4	C5	C6
1	Kluster 1	0.538639 28112965	0.159687 0685688	0.175609 75609756	0.695652 17391304	0.152980 22802713	0.115731 70731707
2	Kluster 2	0.429239 76608187	0.151991 61425577	0.447619 04761905	0.329710 14492754	0.177021 69625247	0.127777 77777778
3	Kluster 3	0.444582 04334365	0.632630 41065483	0.134453 78151261	0.547314 57800511	0.282457 36164288	0.275588 23529412
4	Kluster 3	0.400328 94736842	0.179834 90566038	0.150892 85714286	0.387567 93478261	0.160225 59171598	0.111328 125

Melakukan perulangan dari langkah 3 hingga 5, sampai anggota tiap cluster tidak ada yang berubah. Setelah melakukan proses perhitungan menggunakan centroid ke 9, peneliti menampilkan hasil keseluruhan clustering.

Irfan Wajdi

$$\sqrt{(0.4421 - 0.5386)^2 + (0.0566 - 0.1597)^2 + (0.5429 - 0.1756)^2 + (0.2391 - 0.6957)^2 + (0.1183 - 0.153)^2 + (0.135 - 0.1157)^2}$$

$$= \sqrt{(0.0093 + 0.0106 + 0.1349 + 0.2084 + 0.0012 + 0.0004)}$$

$$= \sqrt{(0.3648)}$$

$$= 0.6040$$

$$\sqrt{(0.4421 - 0.4292)^2 + (0.0566 - 0.152)^2 + (0.5429 - 0.4476)^2 + (0.2391 - 0.3297)^2 + (0.1183 - 0.177)^2 + (0.135 - 0.1278)^2}$$

$$= \sqrt{(0.0002 + 0.0091 + 0.0091 + 0.0082 + 0.0034 + 0.0001)}$$

$$= \sqrt{(0.0300)}$$

$$= 0.1733$$

$$\sqrt{(0.4421 - 0.4446)^2 + (0.0566 - 0.6326)^2 + (0.5429 - 0.1345)^2 + (0.2391 - 0.5473)^2 + (0.1183 - 0.2825)^2 + (0.135 - 0.2756)^2}$$

$$= \sqrt{(0.0000 + 0.3318 + 0.1668 + 0.0950 + 0.0269 + 0.0198)}$$

$$= \sqrt{(0.6403)}$$

$$= 0.8002$$

$$\sqrt{(0.4421 - 0.4003)^2 + (0.0566 - 0.1798)^2 + (0.5429 - 0.1509)^2 + (0.2391 - 0.3876)^2 + (0.1183 - 0.1602)^2 + (0.135 - 0.1113)^2}$$

$$= \sqrt{(0.0017 + 0.0152 + 0.1536 + 0.0220 + 0.0018 + 0.0006)}$$

$$= \sqrt{(0.1949)}$$

$$= 0.4415$$

Yuni Sulaeman

$$\sqrt{(0.3895 - 0.5386)^2 + (0.1509 - 0.1597)^2 + (0.1714 - 0.1756)^2 + (0.3913 - 0.6957)^2 + (0.1657 - 0.153)^2 + (0.125 - 0.1157)^2}$$

$$= \sqrt{(0.0223 + 0.0001 + 0.0000 + 0.0926 + 0.0002 + 0.0001)}$$

$$= \sqrt{(0.1152)}$$

$$= 0.3394$$

$$\sqrt{(0.3895 - 0.4292)^2 + (0.1509 - 0.152)^2 + (0.1714 - 0.4476)^2 + (0.3913 - 0.3297)^2 + (0.1657 - 0.177)^2 + (0.125 - 0.1278)^2}$$

$$= \sqrt{(0.0016 + 0.0000 + 0.0763 + 0.0038 + 0.0001 + 0.0000)}$$

$$= \sqrt{(0.0818)}$$

$$= 0.2860$$

$$\sqrt{(0.3895 - 0.4446)^2 + (0.1509 - 0.6326)^2 + (0.1714 - 0.1345)^2 + (0.3913 - 0.5473)^2 + (0.1657 - 0.2825)^2 + (0.125 - 0.2756)^2}$$

$$\begin{aligned}
 &= \sqrt{(0.0030 + 0.2320 + 0.0014 + 0.0243 + 0.0136 + 0.0227)} \\
 &= \sqrt{(0.2971)} \\
 &= 0.5450 \\
 &\sqrt{(0.3895 - 0.4003)^2 + (0.1509 - 0.1798)^2 + (0.1714 - 0.1509)^2 + (0.3913 - 0.3876)^2 + (0.1657 - 0.1602)^2 + (0.125 - 0.1113)^2} \\
 &= \sqrt{(0.0001 + 0.0008 + 0.0004 + 0.0000 + 0.0000 + 0.0002)} \\
 &= \sqrt{(0.0016)} \\
 &= 0.0401
 \end{aligned}$$

Tabel 11 Hasil Clustering

No	Nama	Chol	TG	HDL	LDL	Cr	BUN	Cluster
1	Wulan Hadinata	4.9	1.6	1	3.2	55	3.4	1
2	Erni Kumolo	4.5	1.2	1.8	4.1	88	6.3	1
3	Wati Prasetyo	4.9	1.6	1	3.2	55	3.4	1
4	Achmad Corbuzier	4.6	0.8	0.9	4.2	55	4.6	1
5	Kusuma Zulkarnain	6.5	1.5	0.9	4.9	67	5.6	1
6	Andi Mahendra	3.9	2.1	1.2	4.2	106	5	1
7	Nugroho Siregar	3.9	1.6	0.9	3.3	79	5.5	1
8	Emma Sutedjo	4.4	1.6	0.8	3	48	3.4	1
9	Bima Anggara	4.6	1.5	0.7	3	123	5.8	1
10	Yulia Mulyadi	9.5	1.7	1.3	2.5	39	3	1
..
140	Maya Zulkarnain	4.8	2.5	1.1	2.7	63	5	4

Dari total sampel penelitian sebanyak 140 pasien berhasil dikelompokkan ke dalam 4 klaster optimal (K = 4). Proses partisi data ini menghasilkan sebaran populasi yang bervariasi di setiap klasternya, mencerminkan adanya heterogenitas karakteristik klinis atau demografis di antara subjek penelitian. Berdasarkan hasil analisis terhadap 140 pasien, populasi didominasi oleh kelompok Kritis/Kronis (Klaster 4) sebesar 45,7% (64 pasien), diikuti kelompok Kelompok Sehat/Normal (Klaster 1) sebesar 30% (42 pasien), serta kelompok Risiko Rendah/Awal (Klaster 2) dan Risiko Tinggi (Klaster 3) yang memiliki proporsi minoritas setara masing-masing sebesar 12,1% (17 pasien).

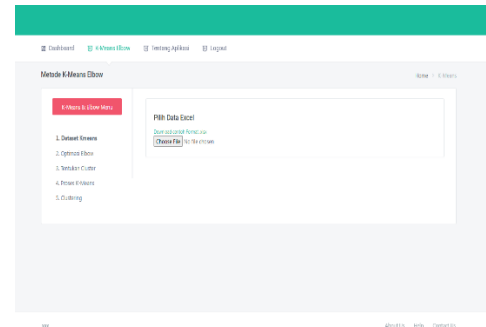


Gambar 3 Grafik Hasil Clustering

Tahapan implementasi merupakan proses pengujian terhadap aplikasi yang dirancang untuk menyesuaikan antara rancangan sistem dan desain aplikasi. Berikut ini beberapa tampilan dari implementasi aplikasi yang peneliti jalankan:

Tampilan Menu K-Means Elbow

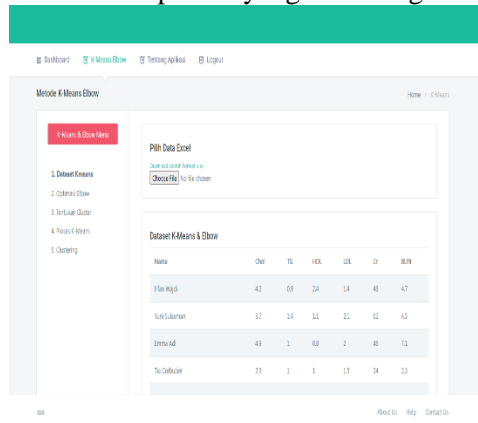
Menu K-Means Elbow digunakan untuk membantu program memahami cara menerapkan teknologi seperti data mining untuk mempelajari dan menghasilkan hasil yang canggih. Ini dapat dilengkapi dengan set data berikutnya yang disebut set validasi dan pengujian. Berikut tampilan dari menu K-Means Elbow.



Gambar 4 Tampilan Menu K-Means Elbow

Tampilan Menu Dataset K-Means

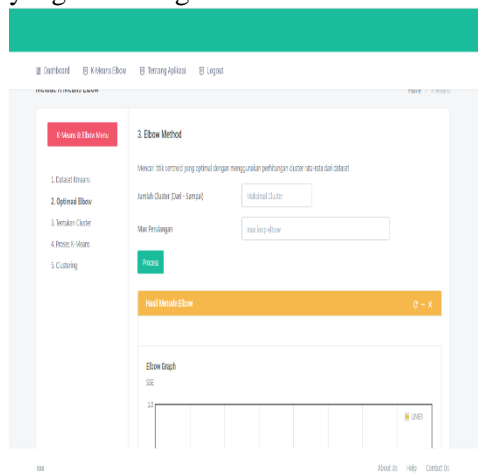
Menu dataset K-means digunakan untuk melakukan penginputan dataset sebelum melakukan pengujian data. Berikut tampilan dari menu dataset K-means dari aplikasi yang dirancang.



Gambar 5 Tampilan Menu Dataset K-Means

Tampilan Menu Optimasi Elbow

Menu Optimasi Elbow digunakan untuk mencari jumlah kluster yang terbaik dalam menampilkan tingkat akurasi yang tepat. Berikut tampilan dari menu Optimasi Elbow yang dirancang.

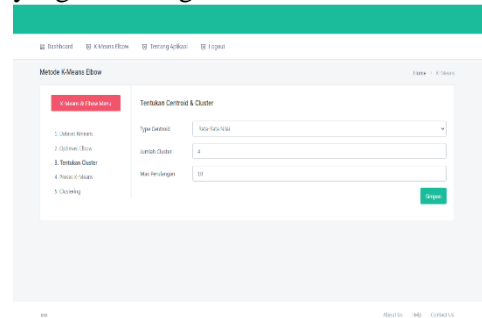


Gambar 6 Tampilan Menu Optimasi Elbow

Tampilan Menu Tentukan Cluster

Menu Tentukan Cluster digunakan untuk menginputkan jumlah kluster yang terbaik untuk proses K-

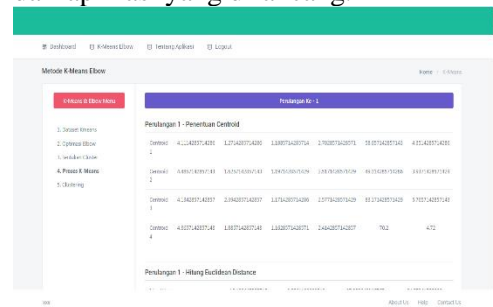
Means. Berikut tampilan dari menu yang dirancang.



Gambar 7 Tampilan Menu Tentukan Cluster

Tampilan Menu Proses K-Means

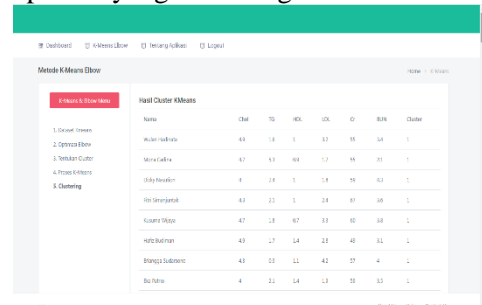
Menu Proses K-Means digunakan untuk proses pengujian dataset. Berikut tampilan dari menu Proses K-Means dari aplikasi yang dirancang.



Gambar 8 Tampilan Menu Proses K-Means

Tampilan Menu Clustering

Menu Clustering digunakan untuk melihat hasil pengelompokan data berdasarkan hasil pengujian. Berikut tampilan dari menu clustering dari aplikasi yang dirancang.



Gambar 9 Tampilan Menu Clustering

SIMPULAN

Dari hasil pembahasan pada sebelumnya yang telah diuraikan, maka peneliti mengambil kesimpulan sebagai berikut:

1. Algoritma K-Means berhasil diterapkan untuk mengelompokkan tingkat risiko penyakit kronis dengan cara mempartisi data klinis darah pasien ke dalam klaster-klaster homogen berdasarkan kedekatan jarak Euclidean (*Euclidean Distance*). Proses iteratif K-Means yang meminimalkan varians di dalam klaster terbukti mampu memisahkan populasi pasien secara tegas ke dalam tingkatan risiko yang terukur, mulai dari Kelompok Sehat (Normal), Risiko Rendah/Awal, Risiko Tinggi, hingga kondisi Kritis/Kronis.
2. Penerapan metode Elbow memberikan peningkatan performa dan validitas yang signifikan pada algoritma K-Means dengan cara memecahkan kelemahan utama K-Means, yaitu penentuan jumlah klaster (K) secara acak. Melalui grafik nilai *Within-Cluster Sum of Squares* (WCSS) atau tingkat inersia dari berbagai variasi nilai K, metode Elbow membantu mendeteksi titik tekukan (elbow) yang paling optimal dalam penelitian ini menghasilkan K=4. Optimalisasi ini memastikan struktur klaster yang terbentuk memiliki kohesi internal yang tinggi (jarak antar-data dalam klaster dekat) dan pemisahan eksternal yang jelas (jarak antar-klaster jauh), sehingga mencegah terjadinya *over-clustering* atau *under-clustering* pada data medis pasien.
3. Aplikasi *clustering* berbasis web berhasil dirancang dan dibangun menggunakan arsitektur sistem yang mengintegrasikan bahasa pemrograman seperti PHP untuk memproses komputasi algoritma K-

Means dan Elbow secara *real-time*. Sistem dirancang dengan alur kerja yang dimulai dari *input* data rekam data darah pasien, proses *preprocessing* (normalisasi data), eksekusi mesin klasterisasi otomatis, hingga penyajian luaran (*output*). Aplikasi ini mampu mentransformasikan data tabular yang kompleks menjadi visualisasi dasbor yang informatif (seperti grafik *donut chart* distribusi pasien dan grafik Elbow), sehingga menghasilkan sebuah sistem yang valid, transparan, dan siap digunakan oleh tenaga medis untuk identifikasi dini risiko penyakit kronis.

DAFTAR PUSTAKA

- Aipina, D., & Witriyono, H. (2022). Pemanfaatan Framework Laravel Dan Framework Bootstrap Pada Pembangunan Aplikasi Penjualan Hijab Berbasis Web. *Jurnal Media Infotama*, 18(1), 36–42.
- Budiarti, Alwendi, Andi Saputa Mandopa, L. (2024). Aplikasi Data Mining Untuk Menentukan Lulusan Mahasiswa Tepat Waktu. *Jurnal Advance Research Informatika*, 3(1), 59–65.
- Damanik, B., Zalukhu, S., Hutagalung, D. M., & Ginting, R. U. (2022). Sistem Informasi Penjualan Minuman Berbasis Web Di Toko Ts Boba Burhanuddin. 4(2), 328–338.
- Dhian Luluh Rohmawati, Endri Ekayanti, R. K. (2022). Pemberdayaan Masyarakat Dalam Melakukan Manajemen Penyakit Kronis Sebagai Langkah Preventif Terjadinya Komplikasi Penyakit Ginjal Kronik. *Jurnal Kreativitas Pengabdian Kepada Masyarakat (Pkm)*, 5(11), 3831–3841.
- Ernawati, A., & Wahyuni, S. (2024). Analisis Data Mining Pola Penggunaan Seluler Dan

- Klasifikasi Perilaku Pengguna Di Berbagai Perangkat Menggunakan Metode C4 . 5. *Bulletin Of Information Technology (Bit)*, 5(4), 162–168.
- Erwan Effendi, Rodika Sima Arif Sagalai, S. R. (2023). Jenis-Jenis Sistem Informasi Dan Model Sistem Informasi. *Jurnal Pendidikan Dan Konseling*, 5(2), 4944–4952.
- Fahzirah, I. (2024). *Pengenalan Sistem Database : Konsep Dasar Dan Manfaatnya Dalam Perusahaan*. 1(4), 673–678.
- Irmayani, D., & Munandar, M. H. (2020). Sistem Informasi Pengelolaan Data Siswa Pada Sma Negeri 02 Bilah Hulu Berbasis Web. *Jurnal Informatika*, 8(2), 65–71.
- Ishak, R. (2024). Optimasi K-Means Pada Clustering Penyakit Ibu Hamil Menggunakan Random Forest Optimization Of K-Means In Disease Clustering Of Pregnant Women Using Random Forest. *Jambura Journal Of Electrical And Electronics Engineering*, 7(1), 41–47.
- Perwitasari, I. D., Hendrawan, J., Panggabean, F. Y., & Raihansyah, M. (2024). *Model Uml Aplikasi Augmented Reality Pengenalan Desa Pertumbuhan*. 13, 1887–1896.