

SENASTIKA Universitas Malikussaleh

IDENTIFIKASI TINGKAT RESIKO KESEHATAN JANTUNG DENGAN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING

Putri Syifa*¹, Sarah Oktia Putri², Putri Almunawarah*³, Munirul Ula⁴

^{1,2,3,4} Teknik Informatika, Universitas Malikussaleh, Lhokseumawe
Email: 1putri.210170280@mhs.unimal.ac.id, 2sarah.210170274@mhs.unimal.ac.id,
3putri.210170281@mhs.unimal.ac.id, 4munirulula@unimal.ac.id

Abstrak

Penyakit jantung termasuk salah satu jenis penyakit yang menjadi masalah serius di seluruh dunia dan dapat mengenai siapa saja, tanpa memperdulikan umur atau *gender*. Beragam faktor, termasuk pola hidup yang buruk, asupan konsumsi makanan tinggi kolesterol, kebiasaan merokok, kondisi lingkungan, dan faktor genetik, berperan dalam meningkatkan risiko terjadinya penyakit jantung. Gejala yang sering muncul antara lain adalah nyeri di dada, hipertensi, serta hasil pemeriksaan seperti *Elektrodiagram* (EKG) yang menunjukkan adanya kelainan. Minimnya data mengenai penyakit jantung menyulitkan upaya prediksi dan diagnosis, khususnya untuk pasien dengan data minoritas, yang dapat menyebabkan terjadinya kesalahan klasifikasi. Penelitian ini memiliki tujuan untuk mengelompokkan risiko penyakit jantung dengan memanfaatkan algoritma *K-Means Clustering*, yang membedakan antara kelompok individu yang menderita penyakit jantung (minoritas) dan kelompok individu yang sehat (mayoritas). Studi menunjukkan bahwa *Algoritma K-Means* menghasilkan 2 *cluster* yaitu, *cluster 0* untuk risiko rendah/sedang, dan *cluster 1* untuk risiko tinggi, dan hasil evaluasi *clustering* sebesar 0,5 menggunakan Metode *Elbow*, dengan tingkat akurasi mencapai 89%. Penelitian ini mengindikasikan bahwa dataset yang tidak seimbang menghasilkan tingkat akurasi kesalahan klasifikasi (*missclassification*) yang 4% lebih rendah dibandingkan dengan dataset yang seimbang, menegaskan pentingnya pengelolaan data untuk meningkatkan prediksi risiko penyakit jantung.

Keywords: *Jantung, k-means clustering, Penyakit, Identifika*

1. PENDAHULUAN

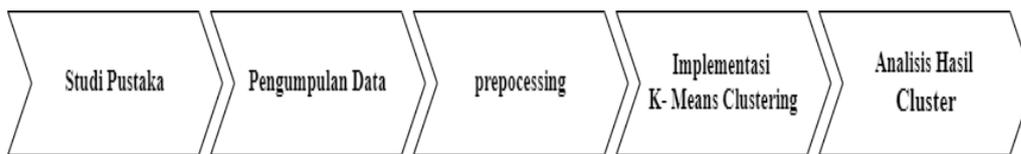
Jantung merupakan peran penting dalam menjaga kelangsungan hidup setiap individu. Penyakit jantung merupakan penyebab utama kematian di seluruh dunia. Menurut *World Health Organization* pada tahun 2020, ditemukan 17,9 juta kematian, di mana 80% di antaranya disebabkan oleh penyakit arteri jantung dan stroke otak [1]. Sebagai organ penting dalam tubuh jantung harus tetap berfungsi secara stabil dan optimal, karena gangguan pada organ ini dapat berakibat fatal, seperti terjadinya gangguan jantung yang disebabkan oleh penyumbatan arteri dan gangguan aliran darah. Jumlah kematian yang besar ini umum terjadi di negara-negara berpenghasilan rendah dan menengah [2]. Penyakit jantung dapat dipengaruhi oleh banyak aspek yang berkaitan dengan kebiasaan hidup, seperti merokok, konsumsi alkohol dan kafein secara berlebihan, stres, serta kurangnya aktivitas fisik. Penyakit ini merupakan salah satu masalah kesehatan yang paling umum di masyarakat, yang dapat menyerang siapa saja tanpa memandang usia, jenis kelamin, atau gaya hidup. Menurut *WHO* (Organisasi Kesehatan Dunia) dan CDC, penyakit jantung adalah penyebab utama kematian di Inggris, Amerika Serikat, Kanada, dan Australia. Di antara populasi orang dewasa, terdapat 26,6 juta orang (11,3%) yang didiagnosis dengan penyakit jantung. [3]. Penanganan penyakit jantung perlu dilakukan tindakan yang cepat tergantung tingkat keparahan yang dialami oleh seorang pasien. Ada berbagai cara untuk mengobati penyakit jantung, di antaranya operasi, terapi radiasi, dan kemoterapi[4]. Masalah utama penyakit jantung adalah sulitnya deteksi dini, sehingga banyak penderita baru menyadari kondisinya ketika penyakit sudah mencapai tahap yang serius. Deteksi penyakit jantung biasanya dilakukan dengan metode konvensional, seperti konsultasi dengan dokter spesialis jantung dan pemeriksaan laboratorium, yang mengakibatkan biaya yang relatif tinggi. [5].

Oleh karena itu, diagnosis dini sangat penting dilakukan. Mendiagnosis penyakit jantung merupakan tantangan besar karena adanya keterkaitan yang kompleks. [6]. Dalam mendeteksi gejala penyakit jantung, penulis menggunakan teknik clustering. Clustering bekerja dengan cara mencari dan mengelompokkan data berdasarkan kesamaan atau kemiripan yang dimilikinya. [7]. K-means adalah salah satu teknik pengelompokan yang bekerja dengan metode *partitioned clustering*. [8]. Dalam clustering, terdapat metode K-Means yang dikenal efisien. Metode ini bertujuan membentuk cluster objek berdasarkan atribut tertentu menjadi k partisi. [9]

Algoritma K-Means merupakan model *Centroid Algoritma K-Means* bekerja dengan membagi data menjadi sejumlah kelompok berdasarkan titik data terhadap masing-masing kelompok. Posisi ini dilakukan secara berulang hingga posisi centroid mencapai stabilitas dan tidak terjadi perubahan didalam pengelompokan. *Algoritma K-Means* dapat diartikan sebagai algoritma pembelajaran yang sederhana, digunakan untuk menyelesaikan masalah pengelompokan dengan tujuan meminimalkan kesalahan ganda. [10]. Mengacu pada banyaknya gejala dan faktor yang terlibat dalam deteksi penyakit jantung, diperlukan pengelompokan untuk mengidentifikasi jenis-jenis penyakit jantung. Dataset dari *kaggle* menunjukkan bahwa terdapat 13 variabel yang berkontribusi terhadap gejala penyakit jantung. Namun, dataset ini bersifat tidak seimbang (*imbalanced*), sehingga diperlukan metode untuk mencapai keseimbangan data, salah satunya adalah *oversampling*. *Oversampling* adalah teknik yang digunakan untuk mengatasi masalah ketidakseimbangan dengan mendistribusikan data secara seimbang melalui replikasi data sintetik dari kelas minoritas secara acak melalui beberapa iterasi [11]. Penelitian ini bertujuan untuk mengelompokkan serta menghasilkan data simulasi pada dataset indikasi penyakit jantung menurut klasterisasi, dengan begitu mampu menghasilkan tingkat akurasi yang sesuai dan optimal.

2. METODE PENELITIAN

Penelitian ini meliputi serangkaian langkah penting untuk mempersiapkan dan merencanakan studi secara komprehensif. Lima tahap yang dilaksanakan pada riset ini mencakup kajian literatur, perekapan data, penerapan metode K-Means clustering, dan analisis hasil klaster. Rincian lengkap mengenai langkah-langkah penelitian bisa diamati di gambar 1.



Gambar 1. Tahap Riset

2.1 Studi Pustaka

Langkah dasar penting dalam melakukan penelitian pendahuluan adalah melakukan studi pustaka yang komprehensif. Ini melibatkan peninjauan yang luas terhadap literatur yang berkenaan dengan Tema studi yang dijalankan [10]. Pada tahap ini riset mengumpulkan dan menganalisis jurnal serta artikel untuk memahami kondisi terkini dan mengidentifikasi bidang yang perlu diteliti lebih lanjut.

2.1 Pengumpulan Data

Penelitian ini memakai data yang tersedia di *Kaggle*, kumpulan dataset penyakit jantung UCI yang diterbitkan pada tahun 2019 dengan jumlah keseluruhan 303 data dapat dilihat di tabel 1. Data ini dikumpulkan yang telah dilakukan survei medis sebelumnya untuk menganalisis faktor-faktor yang mempengaruhi penyakit jantung. Dataset ini memiliki 13 atribut, dan tabel 2 menunjukkan informasi yang dapat diamati.

Tabel 1. Kumpulan Data Penyakit Jantung

ID	age	sex	cp	trtbps	chol	oldpeak	slp	caa	thall	output
1	63	1	3	145	233	2,3	0	0	1	1
2	37	1	2	130	250	3,5	0	0	2	1
3	41	0	1	130	204	1,4	2	0	2	1
4	56	1	1	120	236	0,8	2	0	2	1
...
...
299	57	0	0	140	241	0,2	1	0	3	0
300	45	1	3	110	264	1,2	1	0	3	0
301	68	1	0	144	193	3,4	1	2	3	0
302	57	1	0	130	131	1,2	1	1	3	0
303	57	0	1	130	236	0,0	1	1	2	0

Tabel 2. Deskripsi *Dataset Heart Disease* (Penyakit Jantung)

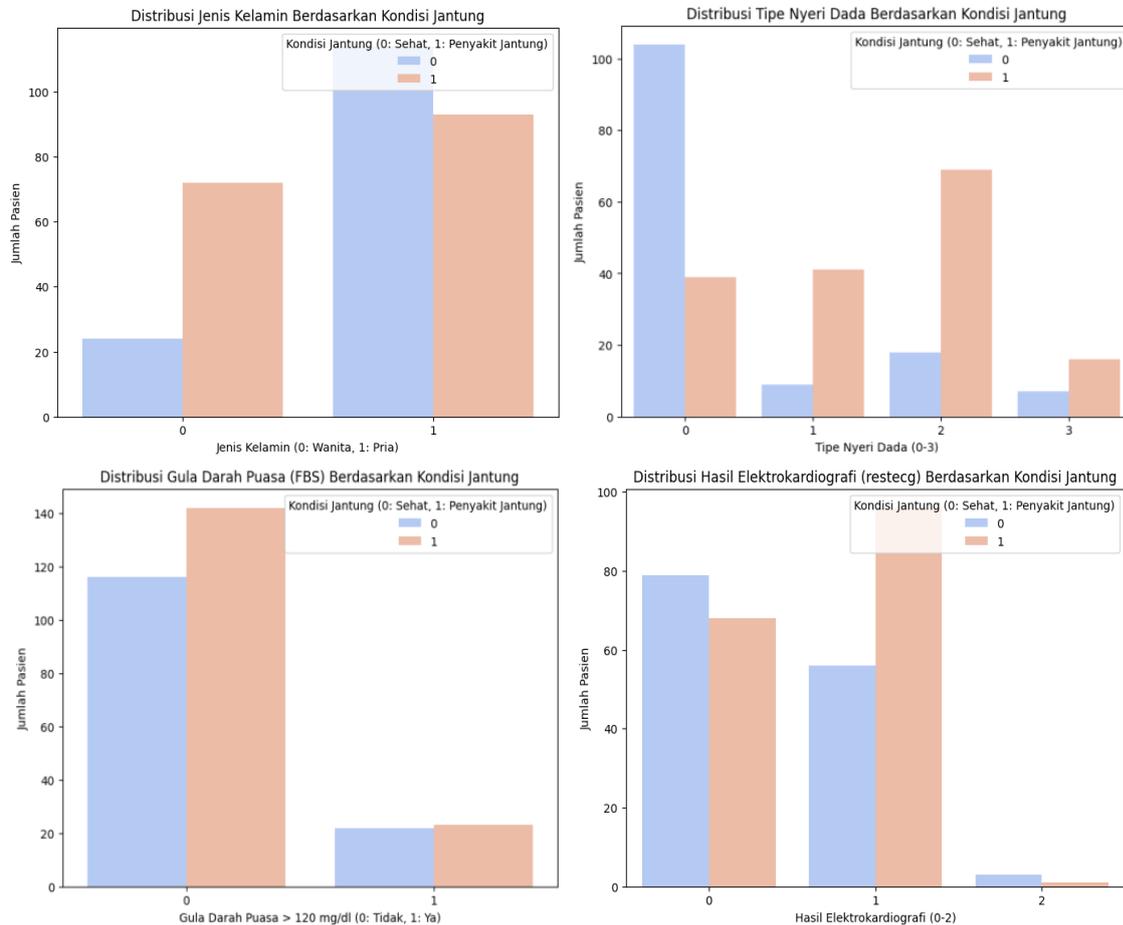
Atribut	Keterangan	Penjelasan
<i>Id</i>	Id Pasien	Kode pasien dari 1-303
<i>Age</i>	umur pasien (tahun)	Minimal = 29, Maksimal =77
<i>Sex</i>	Gender pasien	1 = pria, 0 = wanita
<i>Cp</i>	Tipe nyeri dada	<i>Cp</i> (<i>Chest pain</i>) yaitu klasifikasi nyeri dada yang dialami penderita. Memiliki 4 elemen yaitu : Nilai 1: tidak nyeri dada (<i>no chest pain</i>) Nilai 2: nyeri dada ringan (<i>mildchestpain</i>) Nilai 3: nyeri dada sedang (<i>moderatechestpain</i>) Nilai 4: nyeri dada parah (<i>severechestpain</i>)
<i>Trestbps</i>	Tekanan darah istirahat (mm Hg)	<i>Trestbps</i> (<i>Restingbloodpressure</i>) yaitu tekanan darah pasien saat dalam kondisi istirahat. Rendah < 120, normal = 120, tinggi > 120
<i>Chol</i>	Serum kolesterol (mg/dl)	<i>Chol</i> (Kolesterol) yaitutingkat kolesterol dalam darah pasien. Rendah < 140, normal = 140, tinggi > 140
<i>Fbs</i>	Gula darah puasa >120 mg/dl	<i>Fbs</i> (<i>Fastingbloodsugar</i>) yaitu tingkat kadar gula pasien, dimana FBS untuk 1 melebihi 120 mg/dl, dan 0 jika tidak melebihi atau sama dalam jumlah 120 mg/dl.
<i>Restecg</i>	Hasil elektrokardiografi istirahat	<i>Restingelectrocardiographic</i> memiliki 3 nilai yaitu 0 = normal, 1 = ST-T <i>waveab normality</i> dan 2 = <i>ventricular</i> kiri mengalami <i>hipertrop</i>
<i>Thalach</i>	Denyut jantung maksimum	Tingkat detak jantung maksimum yang dicapai. Jika nilai " <i>thalac</i> " semakin tinggi dapat dianggap sebagai tanda risiko yang lebih tinggi untuk penyakit jantung
<i>Exang</i>	Angina yang dipicu oleh latihan	<i>Exang</i> (<i>Exercise-inducedangina</i>) kondisi dimana pasien akan merasakan nyeri dada.
<i>Oldpeak</i>	Depresi ST	Depresi ST yang disebabkan oleh latihan dibandingkan saat istirahat. Penurunan ST akibat olahraga. Nilai " <i>Oldpeak</i> " yang tinggi dapat dianggap sebagai tanda risiko yang lebih tinggi untuk penyakit jantung
<i>Slp</i>	Kemiringan segmen ST latihan puncak	<i>Slopedari</i> Puncak ST setelah sesi olahraga, Atribut terdiri dari 3 nilai, 0 untuk penurunan, 1 untuk stabil dan 2 untuk meningkat
<i>caa</i>	Jumlah Pembuluh Darah	Nilai 0 = pembuluh darah yang terkena, 1,2,3 = pembuluh darah mengalami penyumbatan
<i>thall</i>	<i>Thallasemia</i>	Nilai 1 =pasien <i>thalasemia</i> normal, 2 = (cacat tetap, yang dapat mempengaruhi fungsi darah) Nilai 3 = kerusakan yang bisa dibalikkan
<i>output</i>		0 = tidak mengalami penyakit jantung, 1 = ada penyakit jantung (atau kondisi tertentu)

Data yang digunakan merupakan data skunder yang dari sumber internet situs Kaggle, *dataset* Penyakit Jantung (*heart disease*).

2.2 Preposeccing

Setelah proses pengumpulan data selesai, langkah selanjutnya adalah melakukan *preprocessing* data. Sebelum diterapkan metode *K-Means Clustering* untuk pengelolaannya. Proses *preprocessing* mencakup penghapusan kolom-kolom yang tidak relevan serta penanganan nilai yang hilang (*missing values*). Pentingnya pengecekan *missing value* dalam analisis data karena hal tersebut dapat membantu mencegah adanya bias dalam penarikan kesimpulan [11]. Metode normalisasi yang akan diterapkan dalam penelitian ini adalah *feature scaling*.

Feature scaling dilakukan untuk membandingkan atau mengintegrasikan data dari berbagai sumber atau variabel yang memiliki rentang nilai yang berbeda. Dengan *feature scaling*, nilainya yang ada akan diubah menjadi rentang yang lebih kecil atau seragam. Proses ini mengkonversi nilai-nilai ke dalam rentang diantara 0 sampai 1.



Gambar 2. Visualisasi data *Exploratory data analysis*

2.4 Implementasi Algoritma *K-Means Clustering*

Proses selanjutnya adalah penerapan algoritma pengelompokan *K-Means*. *K-Means* termasuk salah satu teknik pengelompokan seperti paling *prominent* di dalam ilmu dan teknologi [12]. Tujuan utama dari *K-Means clustering* adalah untuk membagi *dataset* membentuk golongan yang homogen, pada masing-masing golongan memiliki kesamaan internal yang tinggi dan perbedaan yang signifikan antar kelompok [13]. Berikut adalah cara dalam penggunaan *K-Means* untuk segmentasi risiko penyakit jantung:

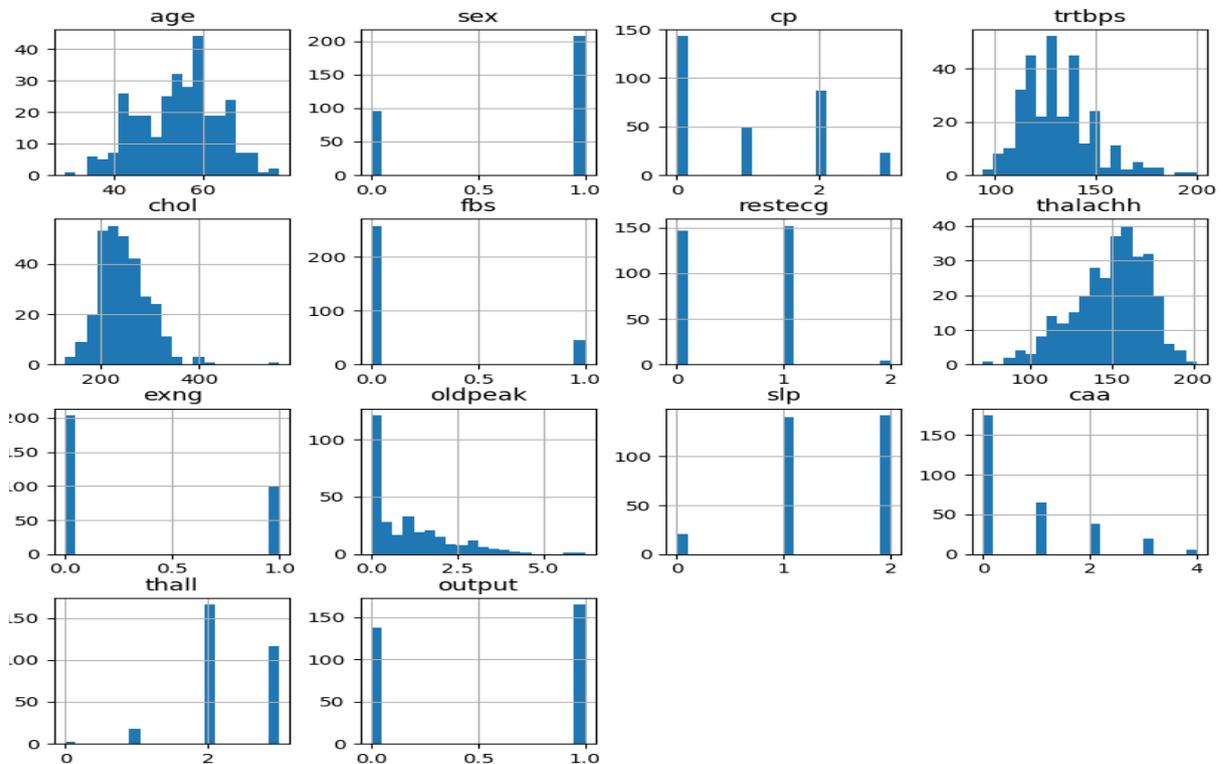
- 1. Inisialisasi Klaster:** Tentukan jumlah klaster *K* yang optimal dengan menggunakan metode seperti *Elbow* atau *Silhouette Score*. Ini bertujuan untuk menemukan jumlah klaster yang paling sesuai untuk membagi data pasien secara efektif, berdasarkan karakteristik risiko penyakit jantung.
- 2. Proses Pengelompokan:** Algoritma *K-Means* membagi pasien menjadi beberapa klaster sesuai dengan nilai fitur kesehatan mereka, seperti usia, tekanan darah, kadar kolesterol, dan detak jantung maksimum. Pasien dengan fitur yang mirip akan ditempatkan dalam satu klaster. *Centroid* atau pusat klaster dihitung berdasarkan rata-rata fitur dari pasien dalam klaster tersebut.
- 3. Iterasi dan Penyempurnaan** Algoritma akan berulang kali memperbarui posisi *centroid* dengan menghitung ulang rata-rata karakteristik kesehatan pada setiap klaster, sambil memindahkan pasien ke klaster baru yang lebih sesuai. Proses ini terus berlangsung hingga tidak ada perubahan signifikan pada posisi *centroid*, yang menandakan bahwa pengelompokan sudah stabil.

2.5 Analisis Hasil Clustering

Tahap analisis hasil *clustering* bertujuan untuk mengidentifikasi kelompok pasien yang perlu diprioritaskan dalam penanganan penyakit jantung. Pemilihan kelompok ini berdasarkan hasil pengelompokan yang dilakukan dengan teknik *clustering*. Data pasien diorganisir berdasarkan karakteristik medis, seperti usia dan riwayat penyakit, serta faktor risiko lainnya. Melalui analisis ini, kelompok dengan risiko tertinggi dapat ditentukan, dan kelompok tersebut menjadi fokus utama dalam keputusan tindakan medis yang akan diambil.

3. HASIL DAN PEMBAHASAN

Berdasarkan hasil tampilan data, peneliti menemukan beberapa faktor yang memengaruhi jumlah pasien dengan penyakit jantung. Faktor-faktor tersebut meliputi jenis nyeri dada (*cp*), yang dialami, tekanan darah saat istirahat (*trtbps*), kadar kolesterol (*chol*), dan gula darah puasa (*fbs*). Hasil *elektrokardiografi* saat waktu rehat (*restecg*), dan senyut jantung tertinggi yang dicapai (*thalachh*) menunjukkan adanya nyeri dada (*angina*), yang dipicu oleh aktivitas fisik, dengan penurunan segmen ST (*slope*) yang terukur. Selain itu, terdapat sejumlah pembuluh darah besar dengan rentang 0-3 yang terlihat melalui *fluoroskopi* (*ca*), dimana 3 menunjukkan kondisi normal, 6 menunjukkan kerusakan permanen, dan 7 menunjukkan kelainan yang dapat kembali (*thal*)



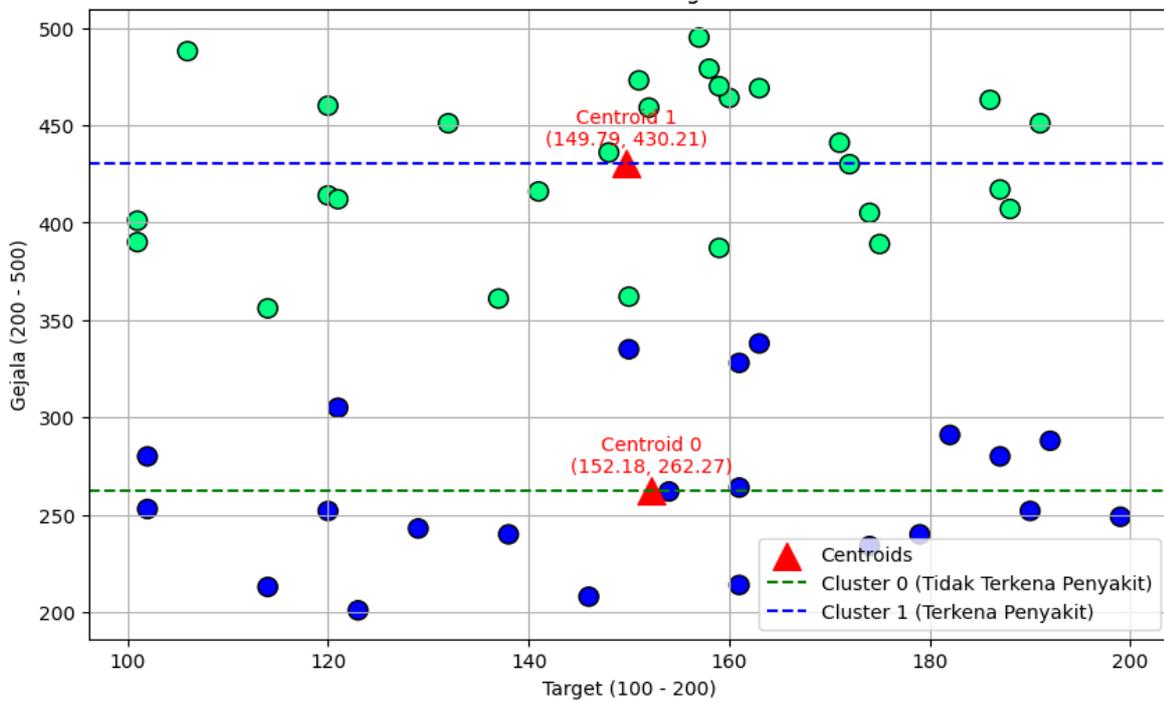
Gambar 3. Ilustrasi visual untuk dataset gejala penyakit jantung

Menurut penggambaran data, terlihat bahwa kumpulan data mengenai gejala penyakit jantung menunjukkan ketidakseimbangan kelas, dengan rasio 4:6. Kelas minoritas terdiri dari 40% dataset, sedangkan kelas mayoritas mencakup 60%. Ketidakseimbangan ini meningkatkan kemungkinan terjadinya kesalahan klasifikasi saat memprediksi diagnosis penyakit. Kesalahan semacam ini dapat menyebabkan penanganan yang tidak tepat bagi pasien, yang dapat memperburuk kondisi, meningkatkan keparahan, bahkan berisiko menyebabkan kematian. Karena itu, diperlukan upaya untuk menangani ketidakseimbangan data guna membuat model prediksi lebih efektif dan akurat sesuai dengan pembagian kelas yang ada.

Penelitian ini mengatasi ketidakseimbangan data dengan menggunakan metode *K-Means Clustering* serta teknik *SMOTE* (*Synthetic Minority Over-sampling Technique*). *SMOTE* menghasilkan data yang disimulasikan mewakili dengan demikian dapat menyederhanakan potensi kesalahan dalam pengelompokan.

Selain itu, dataset dibagi menjadi 2 bagian: data untuk pelatihan (*training*) dan pengujian (*testing*), dengan rasio 1:9. Ini berarti 10% data digunakan untuk pengujian dan 90% untuk pelatihan model.

Proses *clustering* menghasilkan dua atribut, yaitu target dan gejala. Berdasarkan hasil visualisasi, ditemukan adanya dua kelompok (*cluster*). *Cluster 0* menggambarkan individu yang menderita penyakit jantung rendah atau sedang, dengan nilai *centroid* untuk target sebesar 152,18 dan gejala sebesar 262,27. *Cluster* ini diwakili oleh warna biru. Sementara itu, *cluster 1* menunjukkan individu yang menderita penyakit jantung, dengan nilai *centroid* untuk target sebesar 149,79 dan gejala sebesar 430,21, dan diwakili oleh warna merah.



Gambar 4. Ilustrasi Visual Hasil *Clustering* Data

Tabel 3. Kumpulan data dengan 2 *cluster*

ID	cp	Tekanan Darah	Kolestrol	CA	Thal	Target	cluster
1	2	130	250	0	2	1	0
2	1	130	204	0	2	1	0
3	1	120	236	0	2	1	0
4	0	120	354	0	2	1	1
...
...
299	3	110	264	0	3	0	1
300	0	144	193	2	3	0	0
301	0	130	131	1	3	0	0
303	1	130	236	1	2	0	0

Hasil klasterisasi dengan *Algoritma K-Means* menghasilkan 2 *cluster* yaitu :

- **Klaster 0** : pasien yang memiliki resiko rendah atau sedang
- **Klaster 1** : pasien yang memiliki resiko tinggi

Pembahasan

Kluster 0 : Risiko Rendah atau Sedang

Karakteristik : Pasien dalam *cluster* situasi ini menyatakan jika kelompok yang memiliki pola hidup yang lebih baik dan sehat, dengan usia rata-rata antara 35 hingga 45 tahun, BMI yang rata-rata berada dalam kategori sehat, kadar kolesterol yang rendah, serta rutin berolahraga.

Implikasi : Kelompok ini menunjukkan bahwa mereka tidak memerlukan perawatan yang intensif, tetapi tetap perlu pemantauan jangka panjang untuk mencegah peningkatan risiko.

Kluster 1: Risiko Tinggi

Karakteristik : Pasien dalam *cluster* dalam situasi ini menyatakan bahwa kelompok yang memiliki pola hidup yang tidak sehat. Dengan usia rata-rata melebihi 50 tahun, mereka memiliki BMI yang lebih tinggi, kemungkinan berada di atas batas obesitas, serta mempunyai tingkat kolestrol darah serta tekanan aliran darah yang meningkat.

Implikasi : Pasien dalam kluster ini menunjukkan bahwa mereka memerlukan perawatan medis yang lebih serius atau pengobatan yang intensif.

4. DISKUSI

Temuan dari penelitian ini mengindikasikan bahwa penggunaan *Algoritma K-Means Clustering* terbukti efektif dalam mengelompokkan data pasien menurut tingkat risiko kesehatan jantung. Analisis data mengungkapkan bahwa metode ini berhasil mengidentifikasi tiga kelompok utama, yaitu pasien dengan risiko rendah, sedang, dan tinggi. Pengelompokkan ini dilakukan dengan mempertimbangkan variabel-variabel penting, seperti usia, tekanan darah, kadar kolesterol, serta riwayat kesehatan keluarga.

Hasil pengelompokkan ini sesuai dengan penelitian yang telah dilakukan sebelumnya oleh Meida Cahyo Untoro (2023), yang juga mengungkapkan bahwa *Algoritma K-Means* efektif dalam mengelompokkan pasien berdasarkan karakteristik kesehatan jantung. Penelitian tersebut menunjukkan adanya kesamaan dalam pembentukan kelompok, di mana pasien dengan faktor risiko tinggi, seperti hipertensi dan kadar kolesterol yang tinggi, tergolong dalam kategori yang sama. Hal ini menegaskan bahwa faktor-faktor tersebut memiliki dampak yang signifikan terhadap kesehatan jantung.

5. KESIMPULAN

Berdasarkan percobaan yang telah diselesaikan dalam studi ini, disimpulkan bahwa metode klasterisasi dengan *Algoritma K-Means* memberikan wawasan baru terkait pengelompokan tingkat risiko penyakit jantung berdasarkan usia menjadi dua *cluster*. *Cluster 0* mengelompokkan kategori usia dengan risiko penyakit jantung yang tergolong sedang atau rendah, sedangkan *cluster 1* mengelompokkan kategori usia dengan risiko penyakit jantung yang tinggi. Melalui pendekatan yang berbeda ini memberi kesempatan kepada penyedia layanan kesehatan dalam memberikan perawatan yang lebih efisien, yang sesuai dengan tingkatan *cluster* yang alami oleh masing masing pasien serta meningkatkan kualitas hidup pasien melalui pengelolaan penyakit yang lebih individual dan efektif. Diperlukan uji klinis tambahan untuk memastikan bahwa pendekatan ini dapat diimplementasikan secara luas dan memberikan manfaat yang optimal bagi pasien.

DAFTAR PUSTAKA

- [1] Ali, M. M., Paul, B. K., Ahmed, K., Bui, F. M., Quinn, J. M. W., & Moni, M. A. (2021). Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. *Computers in Biology and Medicine*, 136, 104672. <https://doi.org/10.1016/j.compbiomed.2021.104672>
- [2] Shah, D., Patel, S., & Bharti, S. K. (2020). Heart Disease Prediction using Machine Learning Techniques. *SN Computer Science*, 1(6), 345. <https://doi.org/10.1007/s42979-020-00365-y>
- [3] J. Ahmad, A. ul Hasan, T. Naqvi, and T. Mubeen, "A Review on Software Testing and Its Methodology," *Manag. J. Softw. Eng.*, vol. 13, no. 1, pp. 32–38, 2019, doi: 10.26634/jse.13.3.15515.
- [4] E. A. Shams and A. Rizaner, "A novel support vector machine based intrusion detection system for mobile ad hoc networks," *Wirel. Networks*, vol. 24, no. 5, pp. 1821–1829, 2018, doi: 10.1007/s11276-016-1439-0.
- [5] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," *J. Comput. Sci.*, vol. 25, no. 1, pp. 152–160, 2018, doi: 10.1016/j.jocs.2017.03.006.

- [6] Y. I. Kurniawan, A. Rahmawati, N. Chasanah, and A. Hanifa, "Application for determining the modality preference of student learning," in *Journal of Physics: Conference Series*, 2019, vol. 1367, no. 1, pp. 1–11, doi: 10.1088/1742-6596/1367/1/012011.
- [7] Y. Guo, S. Han, Y. Li, C. Zhang, and Y. Bai, "K-Nearest Neighbor combined with guided filter for hyperspectral image classification," in *International Conference On Identification, Information and Knowledge in the Internet of Things*, 2018, pp. 159–165.
- [8] Y. I. Kurniawan, E. Soviana, and I. Yuliana, "Merging Pearson Correlation and TAN-ELR algorithm in recommender system," in *AIP Conference Proceedings*, 2018, vol. 1977, doi: 10.1063/1.5042998.
- [9] M. Sridevi, S. Aishwarya, A. Nidheesha, and D. Bokadia, *Anomaly Detection by Using CFS Subset and Neural Network with WEKA Tools*. Springer Singapore.
- [10] C. Low, "NSL-KDD Dataset," 2015. https://github.com/defcom17/NSL_KDD (accessed Sep. 13, 2019).
- [11] D. Handoko, "Sistem Pendukung Keputusan Seleksi Penentuan Penerima Beasiswa Dengan Metode Simple Additive Weighting (SAW)," Universitas Muhammadiyah Surakarta, 2016.
- [13] Muslimah, V. (2024). Implementing Bayes' Theorem Method in Expert System to Determine Infant Disease. *Khazanah Informatika : Jurnal Ilmu Komputer Dan Informatika*, 10(1), 1–14. <https://doi.org/10.23917/KHIF.V10I1.4837>
- [14] Han, J., & Kang, S. (2023). Optimization of missing value imputation for neural networks. *Information Sciences*, 649, 119668. <https://doi.org/10.1016/j.ins.2023.119668>
- [15] Das, D., Kayal, P., & Maiti, M. (2023). A K-means clustering model for analyzing the Bitcoin extreme value returns. *Decision Analytics Journal*, 6(2022), 100152. <https://doi.org/10.1016/j.dajour.2022.100152>
- [16] Qi, K.-T., Zhang, H.-S., Zheng, Y.-G., Zhang, Y., & Ding, L.-Y. (2023). Stripe segmentation of oceanic internal waves in SAR images based on Gabor transform and K-means clustering. *Oceanologia*, 65(4), 548–555. <https://doi.org/10.1016/j.oceano.2023.06.006>