

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



Artikel Penelitian

Klasifikasi Penyakit Jantung Berbasis Data Rekam Medis Menggunakan Algoritma *Local Mean K-Nearest Neighbor*

Alayda Zaielamy Putri^a, Muhammad Daud^{b,*}, Hafizh Al Kautsar Aidilof^c

^{a,b,c} Universitas Malikussaleh, Lhokseumawe, Aceh, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 13 Januari 2026

Revisi Akhir: 04 Mei 2026

Diterbitkan Online: 11 Mei 2026

KATA KUNCI

Data Mining,
Jantung,
Klasifikasi,
Rekam Medik,
Local Mean K-Nearest Neighbor,
Machine Learning

KORESPONDENSI

E-mail: mdaud@unimal.ac.id*

ABSTRACT

Penyakit jantung merupakan penyebab kematian utama di dunia dengan diagnosis dini yang penting namun sering terkendala akurasi interpretasi data rekam medis kompleks. Algoritma klasifikasi tradisional seperti K-NN memiliki kelemahan dalam menangani noise dan outliers dalam data medis. Penelitian ini bertujuan mengimplementasikan algoritma Local Mean K-Nearest Neighbor untuk mengklasifikasikan penyakit jantung berdasarkan data rekam medis dengan akurasi yang lebih baik. Dataset terdiri dari 403 observasi dengan 10 variabel meliputi jenis kelamin, umur, tekanan darah, heart rate, respiratory rate, hasil elektrokardiogram, kondisi nyeri dada, dan klasifikasi diagnosis. Metode Local Mean K-NN mengadaptasi konsep K-NN tradisional dengan pendekatan local mean calculation untuk mengatasi noise dan outliers. Tahapan penelitian mencakup preprocessing data, feature encoding, feature scaling, hyperparameter tuning, dan evaluasi menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil menunjukkan algoritma Local Mean K-NN dengan nilai K optimal 11 mampu mengklasifikasikan penyakit jantung dengan accuracy 71.60%, precision 69.21%, recall 71.60%, dan F1-score 70.27%. Model menunjukkan performa sangat baik dalam mendeteksi Penyakit Jantung Koroner dengan precision 91.89% dan recall 97.14%. Analisis feature importance mengidentifikasi nyeri dada sebagai indikator terpenting (73.79%), diikuti heart rate (36.40%) dan respiratory rate (25.25%). Penelitian membuktikan efektivitas Local Mean K-NN sebagai clinical decision support tool dalam klasifikasi penyakit kardiovaskular meskipun terdapat tantangan class imbalance pada kelas minoritas.

1. PENDAHULUAN

Data Mining merupakan suatu proses sistematis yang berfokus pada eksplorasi dan analisis data dengan tujuan mengekstraksi atau menyaring informasi dari kumpulan data berskala besar [1]. Proses ini dilakukan melalui serangkaian tahapan yang terstruktur guna memperoleh wawasan yang bernilai [2]. Secara lebih rinci, data mining dapat didefinisikan sebagai seperangkat metode analitis yang bertujuan untuk mengidentifikasi pola atau menemukan pengetahuan tersembunyi dari suatu dataset. Melalui analisis data yang tersimpan dalam basis data, data mining berperan dalam menyelesaikan permasalahan yang kompleks serta memberikan kontribusi signifikan dalam proses pengambilan keputusan. Berbagai teknik dapat diterapkan dalam data mining guna menggali informasi baru, salah satunya adalah

teknik klasifikasi yang memungkinkan pengelompokan data berdasarkan karakteristik tertentu [3].

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia. Berdasarkan laporan *World Health Organization* (WHO), sekitar 17,8 juta jiwa wafat setiap tahunnya yang disebabkan Kerusakan pada sel otot jantung yang berfungsi memompa dan mengalirkan darah ke seluruh tubuh umumnya diawali oleh proses aterosklerosis, yaitu suatu mekanisme kompleks yang menyebabkan terjadinya penumpukan berbagai elemen pada dinding arteri akibat akumulasi lipoprotein plasma. Proses ini berkembang secara bertahap, dimulai dari pembentukan sel busa akibat penumpukan lemak, kemudian berlanjut hingga terbentuknya plak aterosklerotik yang disertai lesi jaringan ikat. Penumpukan plak tersebut dapat menyempit atau menyumbat pembuluh darah arteri, sehingga menghambat

aliran darah dan berpotensi menimbulkan berbagai kejadian klinis [4].

Penelitian ini menerapkan metode kuantitatif dalam menganalisis model klasifikasi *Local Mean K-Nearest Neighbor*. Salah satu fungsi utama dalam kecerdasan buatan adalah proses klasifikasi, yaitu teknik analisis data yang bertujuan untuk mengenali pola-pola spesifik yang dapat membedakan satu kelas data dari kelas data lainnya [5]. Teknik ini digunakan untuk mengkategorikan objek ke dalam kelompok tertentu berdasarkan perilaku dan atribut yang telah didefinisikan sebelumnya [6]. Melalui proses ini, data baru dapat dikelompokkan berdasarkan seperangkat aturan yang diperoleh dari hasil analisis pola yang telah diidentifikasi.

Local Mean K-Nearest Neighbor (LMKNN) adalah sebuah pendekatan dalam klasifikasi yang dikembangkan sebagai penyempurnaan dari algoritma *K-Nearest Neighbor* (KNN) untuk meningkatkan akurasi dalam pengelompokan data. Metode ini dirancang untuk mengurangi pengaruh outlier dalam proses klasifikasi, sehingga meningkatkan akurasi dalam pengelompokan data. Perbedaan mendasar antara LMKNN dan KNN terletak pada cara menentukan parameter k . Pada LMKNN, parameter k dihitung secara terpisah untuk setiap kelas, sedangkan dalam KNN, nilai k merepresentasikan jumlah tetangga terdekat yang dipilih tanpa mempertimbangkan klasifikasi berdasarkan kelas tertentu. Modifikasi ini memungkinkan LMKNN untuk meningkatkan akurasi klasifikasi dengan memperhatikan karakteristik lokal setiap kelas secara lebih spesifik [7].

Permasalahan yang sering ditemui pada aplikasi yang sudah ada dalam klasifikasi penyakit jantung adalah keterbatasan dalam mencapai akurasi yang optimal. Hal ini disebabkan oleh penggunaan algoritma *K-Nearest Neighbor* (KNN) tradisional yang belum mampu menangani outlier dan noise pada data rekam medis, sehingga akurasi hasil klasifikasi menjadi tidak optimal. KNN hanya mempertimbangkan jarak terdekat antar data tanpa memperhitungkan rata-rata lokal, yang menyebabkan hasil klasifikasi kurang stabil.

Penelitian ini bertujuan untuk merancang serta mengembangkan model klasifikasi guna mendeteksi penyakit jantung dengan menerapkan algoritma *Local Mean K-Nearest Neighbor* (LMKNN). Melalui penerapan algoritma ini, diharapkan dapat tercapai peningkatan akurasi prediksi dalam mengidentifikasi dan mengklasifikasikan penyakit jantung, sehingga memberikan hasil yang lebih andal dan efektif dalam mendukung diagnosis medis. Penelitian ini bertujuan untuk mengurangi pengaruh noise dan outlier dalam data rekam medis yang dapat memengaruhi hasil klasifikasi, dengan memanfaatkan pendekatan rata-rata lokal (*local mean*). Dengan demikian, penelitian ini diharapkan dapat menghasilkan metode klasifikasi penyakit jantung yang lebih akurat dan dapat diandalkan dalam mendiagnosis penyakit jantung berdasarkan data rekam medis.

Penelitian sebelumnya telah membahas penggunaan model klasifikasi penyakit jantung, namun, hingga saat ini metode *Local Mean K-Nearest Neighbor* (LMKNN) belum secara spesifik diterapkan dalam penelitian terkait. Sebagaimana dijelaskan dalam penelitian yang dilakukan oleh Akbarollah et al.

berjudul *Penerapan Algoritma K-Nearest Neighbor pada Klasifikasi Penyakit Jantung* [8], metode *K-Nearest Neighbor* (KNN) telah digunakan sebagai pendekatan dalam klasifikasi penyakit jantung. Seiring dengan perkembangan waktu dan semakin beragamnya penerapan teknik klasifikasi untuk berbagai studi kasus, penulis memilih fokus pada klasifikasi penyakit jantung dan mencoba menerapkan metode alternatif, yaitu *Local Mean K-Nearest Neighbor* (LMKNN), untuk mengeksplorasi potensi peningkatan akurasi dan efektivitas dalam analisis tersebut. Dengan menggunakan metode ini untuk mengklasifikasi penyakit jantung, diharapkan dapat mengembangkan model klasifikasi penyakit jantung serta menciptakan model berbasis algoritma *Local Mean K-Nearest Neighbor* (LMKNN).

2. METODE

2.1. Waktu dan Tempat Penelitian

Data yang digunakan dalam penelitian ini bersumber dari penelitian terdahulu yang berjudul "Klasifikasi Penyakit Jantung Menggunakan Metode *Modified K-Nearest Neighbor*" yang dilakukan oleh [3] yang berlangsung di Rumah Sakit Sari Mutiara Lubuk Pakam, Provinsi Sumatera Utara pada bulan Mei 2023 sampai dengan bulan Juli 2023. Penelitian ini sendiri berlangsung pada bulan Juli 2025 hingga selesai, yang berlokasi di Universitas Malikussaleh, Aceh.

2.2. Tahapan Penelitian

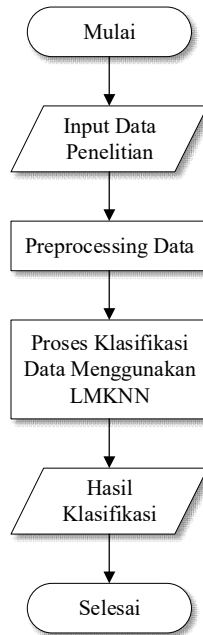
1. Studi literatur, dilakukan untuk mengumpulkan dan mempelajari referensi yang berkaitan dengan penyakit jantung, data rekam medis, serta algoritma *Local Mean K-Nearest Neighbor* sebagai landasan teoritis penelitian.
2. Normalisasi data, bertujuan untuk menyamakan skala nilai setiap fitur sehingga tidak terjadi dominasi atribut tertentu dalam proses perhitungan jarak.
3. Preprocessing data, dilakukan untuk meningkatkan kualitas dataset melalui pembersihan data, penanganan data yang tidak konsisten, serta transformasi data agar siap digunakan dalam proses klasifikasi.
4. Implementasi algoritma *Local Mean KNN*, yaitu tahap penerapan algoritma untuk membangun model klasifikasi penyakit jantung berdasarkan data yang telah diproses.
5. Evaluasi model, dilakukan dengan menggunakan metrik evaluasi untuk mengukur kinerja dan tingkat akurasi model klasifikasi secara objektif.
6. Analisa dan pembahasan, merupakan tahap akhir dimana hasil evaluasi model dianalisis dan diinterpretasikan untuk menilai efektivitas metode yang digunakan sebelum penelitian disimpulkan dan diselesaikan.

2.3. Skema Sistem

Adapun skema sistem yang digunakan pada penelitian ini adalah sebagaimana disajikan pada Gambar 1. Alur proses diawali dengan tahap *Mulai*, kemudian dilanjutkan dengan input data penelitian yang bersumber dari data rekam medis pasien sebagai bahan utama dalam proses klasifikasi. Setelah data dimasukkan, dilakukan tahap *preprocessing* data yang bertujuan untuk

membersihkan, menyesuaikan format, serta menyiapkan data agar sesuai untuk diproses oleh algoritma. Data yang telah melalui tahap *preprocessing* selanjutnya diproses pada tahap klasifikasi menggunakan algoritma *Local Mean K-Nearest Neighbor (LMKNN)* untuk menentukan kelas penyakit jantung

berdasarkan kemiripan karakteristik data. Hasil dari proses klasifikasi kemudian ditampilkan sebagai keluaran sistem berupa hasil diagnosis atau kategori penyakit jantung, dan seluruh rangkaian proses tersebut diakhiri pada tahap *Selesai*.



Gambar 1. Skema Sistem

2.4. Pengertian Jantung dan Penyakit Jantung

Jantung adalah organ esensial dalam sistem tubuh manusia yang berperan utama dalam memompa darah ke seluruh jaringan dan organ. Keberlangsungan fungsi fisiologis tubuh secara optimal sangat bergantung pada kesehatan jantung yang terjaga dengan baik. Apabila kondisi jantung tidak sehat, individu berisiko mengalami berbagai komplikasi kesehatan yang dapat berdampak signifikan terhadap keseimbangan dan kinerja tubuh secara keseluruhan [9].

Jantung adalah organ berotot dengan struktur berongga yang berfungsi sebagai pompa utama dalam sistem *kardiovaskular*, bekerja melalui kontraksi ritmis yang berlangsung secara kontinu. Peran utamanya dalam sirkulasi darah memungkinkan distribusi oksigen dan nutrisi ke seluruh jaringan tubuh, sekaligus mendukung eliminasi produk limbah metabolik. Dari segi anatomi, jantung terletak di dalam *kavum toraks*, sedikit bergeser ke sisi kiri tubuh [10].

Jantung adalah salah satu organ tubuh yang mempunyai empat bagian utama dan berperan dalam memompa darah, baik untuk dikembalikan ke jantung maupun disirkulasikan ke seluruh tubuh. Fungsi ini menjadikan jantung sebagai organ vital dalam mempertahankan kehidupan manusia. Gangguan pada jantung dapat memengaruhi sistem pernapasan, yang pada kondisi tertentu dapat meningkatkan risiko kematian [11].

Penyakit jantung merupakan kondisi di mana organ jantung mengalami gangguan dalam menjalankan fungsinya secara

optimal. Gangguan ini umumnya terjadi akibat kerusakan pada jaringan otot jantung yang disebabkan oleh penyempitan arteri koroner. Faktor risiko penyakit jantung dapat diklasifikasikan menjadi dua kelompok, yaitu faktor yang tidak dapat dimodifikasi dan faktor yang dapat dikendalikan. Faktor yang tidak dapat dimodifikasi mencakup usia, jenis kelamin, serta predisposisi genetik atau riwayat keluarga. Sementara itu, faktor yang dapat dikendalikan meliputi tekanan darah tinggi (hipertensi), kadar kolesterol yang berlebihan, obesitas, diabetes, kurangnya aktivitas fisik, serta konsumsi alkohol dalam jumlah berlebihan [12].

Penyakit Jantung Koroner (PJK) adalah gangguan *kardiovaskular* yang terjadi akibat akumulasi plak di *arteri koroner*, yang berpotensi menyebabkan penyempitan atau obstruksi aliran darah yang membawa oksigen (O_2) ke *miokardium*. Untuk mendukung proses pemulihan dan mencegah komplikasi lebih lanjut, penderita dianjurkan menjalani aktivitas fisik yang sesuai dengan kondisi kesehatan mereka. Salah satu permasalahan utama yang kerap dirasakan oleh individu dengan PJK adalah *angina pectoris* atau nyeri dada, yang tidak hanya berdampak pada kualitas hidup tetapi juga dapat membatasi kapasitas mereka dalam melakukan aktivitas sehari-hari [13].

Gagal jantung merupakan suatu kondisi di mana darah masih dapat kembali ke jantung dalam keadaan normal, namun proses pemompaan darah ke jaringan tubuh untuk mendukung kebutuhan metabolisme tidak dapat dilakukan secara optimal. Maka dalam hal ini memperlihatkan bahwa jantung mengalami penurunan fungsi dalam mengedarkan darah secara efektif,

sehingga dapat berdampak pada berbagai gangguan fisiologis dalam tubuh [14].

Aritmia *supraventrikular* adalah gangguan pada jantung yang menyebabkan detak jantung meningkat secara abnormal, sehingga otot jantung tidak memiliki waktu yang cukup untuk mengendur di antara kontraksi. Kondisi ini dapat mengakibatkan berkurangnya pasokan darah ke berbagai organ tubuh, termasuk otak. Aritmia *supraventrikular* terjadi akibat gangguan yang berasal dari atrium atau nodus atrioventrikular (AV), yang menyebabkan peningkatan frekuensi detak jantung hingga mencapai 140–250 kali per menit [15].

2.5. Data Rekam Medis

Rekam medis merupakan catatan faktual yang mencakup kondisi pasien, riwayat penyakit, serta pengobatan yang telah dan sedang dijalani, yang didokumentasikan oleh tenaga kesehatan yang memberikan pelayanan. Dokumentasi rekam medis memiliki peran penting dalam mencatat temuan dan observasi terkait riwayat kesehatan pasien, baik di masa lalu maupun saat ini, termasuk hasil pemeriksaan, tes medis, asuhan klinis, serta implementasi rencana perawatan oleh tenaga medis. Selain itu, rekam medis juga berfungsi sebagai representasi menyeluruh dari layanan yang diberikan kepada pasien serta aspek pelayanan kesehatan yang mendukungnya [16].

2.6. Data Mining

Data mining merupakan suatu pendekatan analitis dalam mengolah data berukuran besar untuk mengidentifikasi pola tersembunyi, tren, atau keterkaitan yang signifikan dalam suatu himpunan data [17]. Proses ini memanfaatkan berbagai sumber informasi, termasuk sistem basis data, gudang data, serta *infrastruktur* penyimpanan lainnya. Sebagai bidang yang bersifat *interdisipliner*, data mining memiliki keterkaitan erat dengan teknologi basis data, manajemen informasi, statistika, kecerdasan buatan, pembelajaran mesin, dan komputasi canggih [18].

Pada dasarnya, *data mining* berfungsi sebagai metode untuk mengekstraksi informasi yang bernilai dan memperoleh wawasan baru dari data dalam jumlah besar melalui penerapan teknik statistik, matematis, serta algoritma kecerdasan buatan dan pembelajaran mesin [19]. Bidang ini memiliki peran penting dalam membantu pengambilan keputusan berbasis data melalui analisis mendalam dan pengenalan pola yang tersembunyi [20].

Selain itu, *data mining* dapat didefinisikan sebagai proses yang memanfaatkan algoritma pembelajaran mesin guna mengeksplorasi, menganalisis, dan secara otomatis mengekstraksi informasi yang bermakna dari data yang tersedia, sehingga memungkinkan pemanfaatan data secara lebih optimal dalam berbagai bidang [21].

2.7. Klasifikasi Data Mining

Klasifikasi merupakan teknik esensial dalam *data mining* yang bertujuan untuk mengelompokkan objek, konsep, dokumen, atau entitas lain ke dalam kelas tertentu berdasarkan karakteristik yang serupa [22]. Proses ini menitikberatkan pada pembentukan model

atau fungsi yang mampu mengidentifikasi serta membedakan berbagai kategori data secara sistematis. Model yang dihasilkan kemudian dapat digunakan untuk menentukan kelas suatu objek yang belum diketahui dengan menganalisis atribut yang memiliki keterkaitan relevan [23]. Teknik ini mampu mengklarifikasi data baru dengan menggunakan hasilnya untuk memberikan sejumlah aturan [24].

2.8. Algoritma Local Mean K-Nearest Neighbor

Metode *Local Mean K-Nearest Neighbor* (LMKNN) adalah perluasan dari metode KNN. *Local Mean K-Nearest Neighbor* adalah metode klasifikasi nonparametrik yang sederhana, efektif, dan tahan terhadap gangguan. Metode ini juga terbukti mampu meningkatkan performa klasifikasi dan mengurangi pengaruh dari outlier, terutama dalam kondisi ukuran sampel kecil. Proses *Local Mean K-Nearest Neighbor* dapat dijelaskan sebagai berikut:

1. Menetapkan nilai K sebagai parameter utama dalam proses klasifikasi.
2. Mengukur jarak Euclidean dengan menghitung selisih antara data uji dan setiap data latih menggunakan rumus Euclidean berikut:

$$D(x, y) = \|x - y\| = \sqrt{\sum_{j=1}^N (x_j - y_j)^2} \quad (1)$$

Keterangan :

- X : Vektor data uji (test data) dengan N fitur.
 Y : Vektor data latih (training data) dengan N fitur.
 x_j, y_j : Nilai fitur ke j dari data uji dan data latih.
 N : Jumlah total fitur dalam dataset.

Rumus ini mengukur seberapa jauh data uji (x) dengan data latih (y). Dalam LMKNN, jarak ini digunakan untuk mengurutkan data latih dan menemukan tetangga terdekat untuk setiap kelas.

3. Menyusun data dalam urutan berdasarkan jarak, dimulai dari yang paling kecil hingga yang terbesar, sesuai dengan nilai k untuk setiap kategori data.
4. Menghitung rata-rata lokal dari k tetangga terdekat dalam setiap kelas dengan persamaan berikut:

$$m_{k,j}^v = \frac{1}{k} \sum_{i=1}^k v_{i,j} \dots \dots \dots (2)$$

Keterangan :

- $m_{k,j}^v$: Vektor rata-rata lokal dari kelas j .
 k : Jumlah tetangga terdekat yang dipilih untuk setiap kelas.
 $v_{i,j}$: Vektor data latih ke i dari kelas j .

Vektor rata-rata lokal ($m_{k,j}^v$) merepresentasikan pusat dari k tetangga terdekat untuk kelas j . Penggunaan rata-rata lokal membantu algoritma LMKNN lebih tahan terhadap gangguan dari outlier, karena nilai rata-rata lebih stabil dibandingkan hanya menggunakan satu titik data (seperti pada k -NN biasa).

5. Menentukan kelas data uji dengan menghitung jarak terdekat ke vektor rata-rata lokal dari setiap kelas data menggunakan persamaan berikut:

$$wc = \operatorname{argmin}_j d(x, m_{k,j}^v), \quad j = 1, 2, \dots, M \quad (3)$$

Keterangan :

- Wc : Kelas prediksi untuk data uji (x).
 argmin_j : Memilih indeks j dari kelas dengan jarak terkecil.
 $d(x, m_{k,j}^v)$: Jarak antara data uji (x) dengan vektor rata-rata lokal $m_{k,j}^v$ dari kelas j .
 M : Jumlah total kelas dalam dataset.

Data uji (x) akan diklasifikasikan ke kelas dengan jarak terkecil ke rata-rata lokalnya ($m_{k,j}^v$). Proses ini lebih efektif dibandingkan hanya mengandalkan tetangga terdekat tunggal, karena mempertimbangkan distribusi lokal dari data latih dalam setiap kelas.

3. HASIL

Merujuk pada rumusan masalah yang telah dijelaskan sebelumnya, penelitian ini bertujuan untuk menerapkan metode *Local Mean K-Nearest Neighbor* dalam mengklasifikasikan penyakit jantung berdasarkan data rekam medis pasien. Sistem yang dikembangkan berbasis pembelajaran mesin menggunakan pendekatan *supervised learning*, yang dirancang untuk memberikan klasifikasi yang akurat mengenai keberadaan penyakit jantung berdasarkan parameter klinis dan variabel medis terkait lainnya. Dalam proses pengolahan data, dilakukan pembersihan dan *preprocessing* data rekam medis guna meningkatkan konsistensi dan akurasi hasil klasifikasi. Dengan menggunakan algoritma *Local Mean K-NN*, diharapkan penelitian ini mampu menghasilkan model klasifikasi yang efektif dalam membantu tenaga medis melakukan diagnosis penyakit jantung secara lebih optimal dan tepat sasaran.

Selain itu, penelitian ini bertujuan untuk mengevaluasi sejauh mana metode *Local Mean K-Nearest Neighbor* dapat memberikan hasil klasifikasi yang akurat dalam konteks data rekam medis penyakit jantung. Evaluasi performa algoritma dilakukan dengan mengukur metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*, yang digunakan untuk menilai kualitas dan keandalan model klasifikasi yang dihasilkan. Metrik ini memberikan gambaran mengenai tingkat keberhasilan sistem dalam mengklasifikasikan keberadaan penyakit jantung sehingga dapat membantu dalam pengambilan keputusan diagnosis yang lebih tepat oleh tenaga medis dan *stakeholder* kesehatan. Melalui pendekatan evaluasi komprehensif ini, penelitian ini bertujuan untuk menghasilkan rekomendasi implementasi algoritma yang dapat memberikan hasil klasifikasi penyakit jantung secara akurat dan memberikan dasar bagi sistem pendukung keputusan medis yang lebih efektif dalam bidang kardiologi.

3.1. Analisa Masalah

Mengklasifikasikan penyakit jantung berdasarkan data rekam medis merupakan tantangan yang kompleks dalam bidang *healthcare*. Diagnosis penyakit jantung dipengaruhi oleh berbagai faktor medis, seperti usia pasien, jenis kelamin, tekanan darah, kadar kolesterol, hasil elektrokardiogram (EKG), *maximum heart rate*, riwayat nyeri dada, dan faktor klinis lainnya seperti tingkat gula darah dan riwayat keluarga. Proses klasifikasi ini memerlukan pemahaman yang mendalam untuk mengidentifikasi pola dan hubungan antar parameter medis, sehingga dapat membantu tenaga medis dalam melakukan diagnosis yang lebih akurat dan mencegah kesalahan dalam penentuan kondisi kesehatan jantung pasien. Namun, pendekatan konvensional yang mengandalkan diagnosis manual atau metode sederhana seringkali kurang akurat dan tidak mampu menangani kompleksitas data medis yang ada, sehingga menyulitkan dalam pengambilan keputusan diagnosis yang tepat waktu.

3.2. Analisa Deskriptif Data

Data penelitian yang digunakan dalam analisis ini mencakup informasi rekam medis pasien untuk klasifikasi penyakit jantung yang dikumpulkan dari berbagai sumber medis, dengan total 403 observasi yang terdiri dari 10 variabel utama. Variabel-variabel tersebut meliputi nomor urut pasien, jenis kelamin, umur dalam satuan tahun, tekanan darah, *heart rate* (HR) dalam satuan *beats per minute*, *respiratory rate* (RR) dalam satuan napas per menit, hasil elektrokardiogram (EKG), klasifikasi diagnosis, kondisi nyeri dada, dan sumber data rekam medis. Data ini memberikan gambaran komprehensif mengenai karakteristik klinis pasien yang mencakup parameter vital, hasil pemeriksaan diagnostik, dan manifestasi gejala yang relevan untuk deteksi penyakit jantung.

NO	JENIS KELAMIN	UMUR	TEKANAN DARAH	HR	RR	EKG	NYERI DADA	SUMBER	KLASIFIKASI	
0	1	LK	63	145/94	150	22	T Wave abnormal	Ada	UCI Repository	Penyakit Jantung Koroner
1	2	LK	37	130/84	187	19	ST elevasi	Ada	UCI Repository	Penyakit Jantung Koroner
2	3	PR	41	130/84	172	20	Normal	Tidak ada	UCI Repository	Aritmia
3	4	LK	56	120/78	178	22	ST depresi	Tidak ada	UCI Repository	Aritmia
4	5	PR	57	120/78	163	18	ST depresi	Tidak ada	UCI Repository	Aritmia
...
398	399	PR	72	140/90	84	19	Atrial Fibrilasi	Ada	Rumah Sakit	Penyakit Jantung Koroner
399	400	LK	73	140/90	91	20	Atrial Flutter	Tidak ada	Rumah Sakit	Aritmia
400	401	LK	74	130/80	88	19	Atrial Fibrilasi	Ada	Rumah Sakit	Penyakit Jantung Koroner
401	402	PR	75	160/100	83	21	Atrial Fibrilasi	Ada	Rumah Sakit	Penyakit Jantung Koroner
402	403	PR	78	160/100	92	20	Atrial Fibrilasi	Ada	Rumah Sakit	Penyakit Jantung Koroner

403 rows × 10 columns

Gambar 2. Dataset Penelitian

Dataset yang terlihat pada Gambar 2 ini menyajikan variasi yang cukup representatif dari populasi pasien dengan berbagai kondisi kesehatan jantung, mulai dari pasien normal hingga pasien dengan indikasi penyakit jantung. Keberagaman data meliputi distribusi jenis kelamin, rentang usia yang bervariasi, spektrum tekanan darah dari normal hingga hipertensi, variasi *heart rate* dan *respiratory rate*, serta berbagai pola hasil EKG yang mencerminkan kondisi elektrofisiologi jantung yang berbeda. Struktur data ini memungkinkan algoritma *Local Mean K-Nearest Neighbor* untuk melakukan pembelajaran pola yang komprehensif dalam mengidentifikasi karakteristik pasien dengan penyakit jantung berdasarkan kemiripan profil medis antar pasien.

3.3. Analisa Pengolahan Dataset

Sub-bab ini menjelaskan secara komprehensif seluruh tahapan analisis dan pengolahan data yang telah dilakukan dalam penelitian ini. Seluruh proses komputasi dan analisis data dilaksanakan menggunakan perangkat lunak *machine learning* yang memungkinkan visualisasi interaktif dan eksekusi perhitungan secara fleksibel. Tujuan utama dari tahapan ini adalah untuk mempersiapkan data rekam medis yang telah terkumpul agar siap digunakan dalam proses klasifikasi, mengidentifikasi pola dan hubungan antar parameter medis, serta menerapkan algoritma *Local Mean K-Nearest Neighbor* untuk membangun model klasifikasi penyakit jantung yang akurat dan reliabel.

1. Import Library

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
import plotly.figure_factory as ff
```

```

from plotly.subplots import make_subplots
import plotly.offline as pyo
pyo.init_notebook_mode()

from sklearn.model_selection import train_test_split,
cross_val_score, GridSearchCV
from sklearn.preprocessing import LabelEncoder,
StandardScaler, OneHotEncoder
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score,
classification_report, confusion_matrix
from sklearn.metrics import precision_score, recall_score,
f1_score

import pickle
import joblib
import json
import os
import warnings
warnings.filterwarnings('ignore')

```

Tahapan awal dalam pengolahan data dimulai dengan mengimpor seluruh *library* dan modul yang diperlukan untuk mendukung proses analisis dan klasifikasi data rekam medis. *Library* pandas dan numpy digunakan sebagai fondasi utama untuk manipulasi dan pengolahan data, sementara matplotlib, seaborn, dan plotly (dengan berbagai modulnya) diimplementasikan untuk visualisasi data baik statis maupun interaktif. Komponen *machine learning* yang menjadi inti penelitian direpresentasikan melalui modul-modul scikit-learn termasuk *train_test_split* untuk pembagian dataset, *KNeighborsClassifier* untuk algoritma klasifikasi utama, berbagai *preprocessing tools* (*LabelEncoder*, *StandardScaler*, *OneHotEncoder*), serta metrik evaluasi seperti *accuracy_score*, *precision_score*, *recall_score*, dan *f1_score*. Selain itu, *library* pickle dan joblib digunakan untuk penyimpanan model, json dan os untuk operasi *file system*, dan konfigurasi *warnings.filterwarnings('ignore')* diterapkan untuk menjaga kebersihan output selama proses analisis.

2. Preprocessing Data

1. Data Cleaning

```

df_clean = df.copy()
columns_to_drop = ['NO', 'SUMBER'] # Drop kolom tidak
relevan
df_clean = df_clean.drop(columns=columns_to_drop,
errors='ignore')

```

2. Feature-Target Separation

```

X = df_clean.drop('KLASIFIKASI', axis=1)
y = df_clean['KLASIFIKASI']

```

3. Feature Classification

```

categorical_features = ['JENIS KELAMIN', 'TEKANAN
DARAH', 'EKG', 'NYERI DADA']
numerical_features = ['UMUR', 'HR', 'RR']

```

4. Label Encoding

```

label_encoders = {}
for col in categorical_features:
    le = LabelEncoder()

```

```

X_encoded[col] =
le.fit_transform(X_encoded[col].astype(str))
label_encoders[col] = le

```

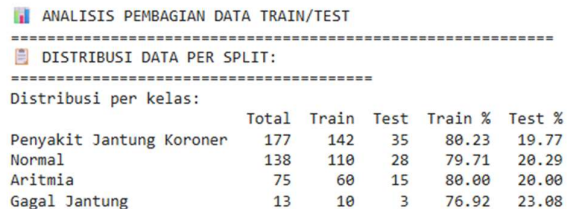
5. Feature Scaling & Data Split

```

scaler = StandardScaler()
X_scaled = scaler.fit_transform(X_encoded)
X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y_encoded, test_size=0.2, random_state=42,
stratify=y_encoded
)

```

Tahapan *preprocessing* data dilakukan untuk memastikan data rekam medis siap digunakan dalam algoritma *Local Mean K-Nearest Neighbor*. Proses diawali dengan *data cleaning* melalui penghapusan kolom yang tidak relevan seperti nomor urut dan sumber data, kemudian dilanjutkan dengan pemisahan antara *features* dan *target* klasifikasi. Fitur diklasifikasikan berdasarkan tipe data menjadi fitur kategorikal dan numerik, dimana fitur kategorikal dikonversi ke bentuk numerik menggunakan *Label Encoding* agar dapat diproses oleh algoritma berbasis jarak. Selanjutnya dilakukan *feature scaling* menggunakan *StandardScaler* untuk menyamakan skala seluruh fitur dan mencegah dominasi nilai tertentu dalam perhitungan jarak. Dataset kemudian dibagi menjadi data latih dan data uji dengan proporsi 80:20 menggunakan *stratified sampling* untuk menjaga keseimbangan kelas. Seluruh komponen *preprocessing* disimpan untuk menjaga konsistensi pada proses prediksi selanjutnya, sehingga data yang dihasilkan bersih, terstruktur, dan siap digunakan dalam pelatihan serta evaluasi model klasifikasi penyakit jantung.



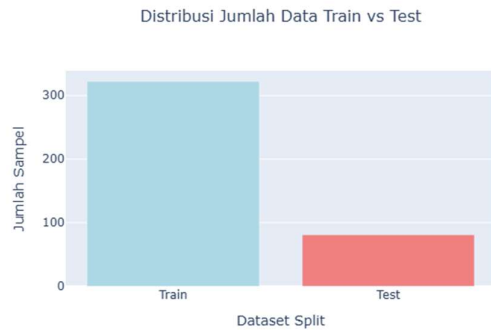
ANALISIS PEMBAGIAN DATA TRAIN/TEST					
DISTRIBUSI DATA PER SPLIT:					
Distribusi per kelas:					
	Total	Train	Test	Train %	Test %
Penyakit Jantung Koroner	177	142	35	80.23	19.77
Normal	138	110	28	79.71	20.29
Aritmia	75	60	15	80.00	20.00
Gagal Jantung	13	10	3	76.92	23.08

Gambar 3. Analisis Pembagian Dataset

Pembagian dataset yang terlihat pada Gambar 3 menggunakan metode *stratified sampling* dengan rasio 80:20 berhasil mempertahankan distribusi kelas yang proporsional antara *training set* dan *testing set*. Dari total 403 observasi, terdapat 322 data untuk pelatihan dan 81 data untuk pengujian model. Distribusi kelas menunjukkan bahwa kelas "Penyakit Jantung Koroner" memiliki jumlah sampel terbanyak dengan 177 observasi (142 untuk *training* dan 35 untuk *testing*), diikuti oleh kelas "Normal" dengan 138 observasi (110 untuk *training* dan 28 untuk *testing*), kelas "Aritmia" dengan 75 observasi (60 untuk *training* dan 15 untuk *testing*), dan kelas "Gagal Jantung" dengan 13 observasi (10 untuk *training* dan 3 untuk *testing*).

Persentase distribusi menunjukkan konsistensi yang baik antara *training set* dan *testing set*, dimana selisih persentase untuk setiap kelas berkisar antara 0.08% hingga 3.85%. Hal ini mengindikasikan bahwa *stratified sampling* berhasil menjaga representativitas data dalam kedua subset, sehingga model yang dilatih pada *training set* dapat dievaluasi secara objektif pada

testing set. Meskipun terdapat ketidakseimbangan kelas (*class imbalance*) yang cukup signifikan, terutama pada kelas "Gagal Jantung" yang hanya memiliki 13 observasi, distribusi ini mencerminkan kondisi riil dalam data medis dimana prevalensi penyakit tertentu memang berbeda-beda dalam populasi.



Gambar 4. Grafik Pembagian Dataset

Grafik *bar chart* pada Gambar 4 menunjukkan distribusi jumlah sampel antara *training set* dan *testing set* setelah pembagian dataset menggunakan rasio 80:20. Visualisasi ini memperlihatkan bahwa *training set* (ditampilkan dalam warna biru) memiliki 322 observasi, sementara *testing set* (ditampilkan dalam warna merah) memiliki 81 observasi. Proporsi pembagian ini mengikuti praktik standar dalam *machine learning* dimana mayoritas data dialokasikan untuk pelatihan model guna memastikan algoritma *Local Mean K-Nearest Neighbor* memiliki cukup data untuk mempelajari pola klasifikasi yang akurat.

Rasio 80:20 yang diterapkan memberikan keseimbangan optimal antara ketersediaan data untuk pelatihan dan validasi objektif model. *Training set* yang berukuran lebih besar memungkinkan algoritma K-NN untuk membangun representasi yang komprehensif dari berbagai karakteristik pasien dengan penyakit jantung, sementara *testing set* yang terpisah menyediakan evaluasi yang tidak bias terhadap kemampuan generalisasi model. Visualisasi ini menegaskan bahwa pembagian data telah dilakukan dengan tepat dan siap untuk tahapan pelatihan serta evaluasi model klasifikasi.

3. Implementasi Local Mean KNN

Implementasi algoritma *Local Mean K-Nearest Neighbor* dilakukan melalui pembuatan kelas khusus yang mengadaptasi

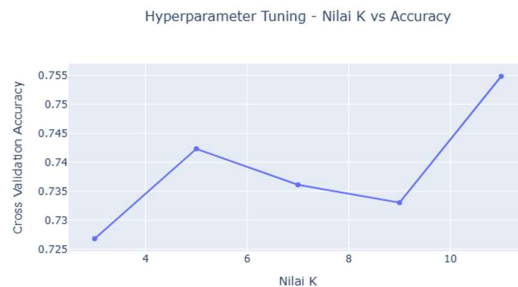
konsep K-NN tradisional dengan pendekatan *local mean calculation*. Algoritma ini bekerja dengan menghitung rata-rata lokal (*local mean*) dari K tetangga terdekat untuk setiap kelas, kemudian melakukan klasifikasi berdasarkan jarak terdekat ke *local mean* tersebut. Pendekatan ini memberikan keunggulan dibandingkan K-NN konvensional karena mampu menangani *noise* dan *outliers* dengan lebih baik melalui penggunaan representasi rata-rata dari tetangga sekelas.

Struktur kelas *LocalMeanKNN* terdiri dari beberapa metode utama yang mencakup *fit()* untuk pelatihan model, *get_k_nearest_neighbors()* untuk mencari K tetangga terdekat menggunakan jarak Euclidean, dan *calculate_local_means()* untuk menghitung rata-rata lokal setiap kelas berdasarkan tetangga yang ditemukan. Metode *predict_single()* melakukan prediksi untuk satu sampel dengan menghitung jarak ke setiap *local mean* dan memilih kelas dengan jarak terkecil, sementara *predict()* dan *predict_proba()* menyediakan fungsionalitas prediksi untuk multiple sampel dan estimasi probabilitas berdasarkan *inverse distance weighting*.

Algoritma ini menggunakan parameter K=5 sebagai default untuk menentukan jumlah tetangga yang dipertimbangkan, dengan fleksibilitas untuk menyesuaikan nilai K sesuai karakteristik dataset. Implementasi juga mencakup penanganan kasus khusus dimana tidak ditemukan tetangga sekelas dalam K tetangga terdekat, dengan menggunakan *global mean* kelas sebagai fallback. Metode *predict_proba()* mengkonversi jarak ke probabilitas menggunakan *inverse distance weighting* dengan *smoothing factor* untuk menghindari pembagian dengan nol, memberikan output probabilitas yang dapat digunakan untuk analisis kepercayaan prediksi dan *threshold tuning*.

4. Training & Evaluasi Model

Grafik *line chart* pada Gambar 6 menunjukkan hasil *hyperparameter tuning* untuk menentukan nilai K optimal dalam algoritma *Local Mean K-Nearest Neighbor*. Visualisasi ini memperlihatkan hubungan antara nilai K (3, 5, 7, 9, 11) dengan *cross-validation accuracy* yang diperoleh dari setiap konfigurasi. Pola yang terlihat menunjukkan bahwa performa model mengalami fluktuasi seiring dengan perubahan nilai K, dimana *accuracy* dimulai dari 0.727 pada K=3, mencapai puncak pada K=5 dengan *accuracy* 0.742, kemudian mengalami penurunan bertahap hingga K=9 dengan *accuracy* 0.733, dan melonjak signifikan pada K=11 dengan *accuracy* tertinggi sebesar 0.755.

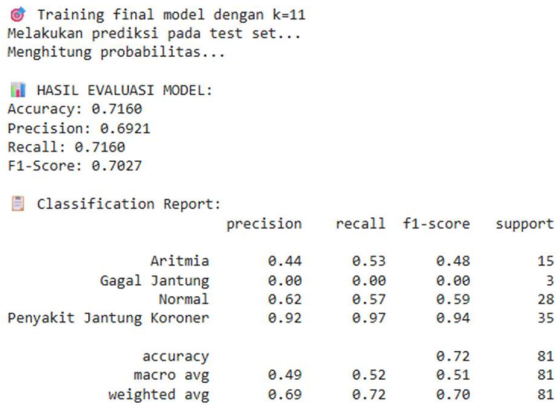


Gambar 5. Grafik Analisis *Hyperparameter Tuning*

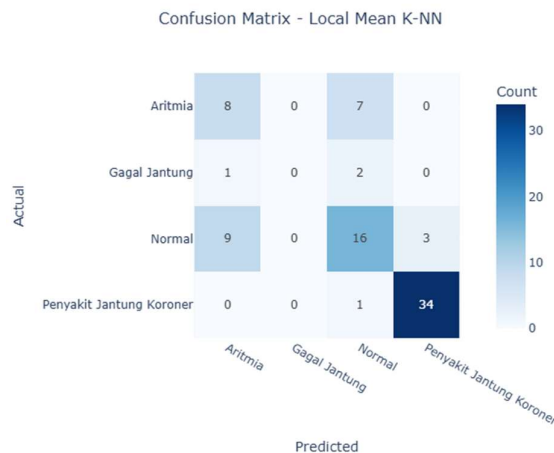
Visualisasi pada Gambar 5 ini memvalidasi pentingnya proses *hyperparameter tuning* dalam optimasi model *machine learning*,

karena perbedaan nilai K dapat memberikan dampak yang signifikan terhadap performa klasifikasi. Peningkatan *accuracy*

dari 0.727 menjadi 0.755 (sekitar 2.8%) antara $K=3$ dan $K=11$ menunjukkan bahwa optimasi parameter dapat memberikan kontribusi yang meaningful terhadap kualitas prediksi model. Hasil ini menjadi dasar pemilihan $K=11$ sebagai parameter optimal untuk model final yang digunakan dalam klasifikasi penyakit jantung.



Gambar 6. Hasil Evaluasi Model *Local Mean KNN*



Gambar 7. Confusion Matrix

Confusion matrix pada Gambar 7 menunjukkan bahwa model *Local Mean K-NN* memiliki kinerja yang sangat baik dalam mengklasifikasikan Penyakit Jantung Koroner, dengan 34 dari 35 kasus diprediksi secara benar (97,1% *recall*), sementara hanya satu kasus mengalami kesalahan klasifikasi. Kinerja yang cukup baik juga ditunjukkan pada kelas Normal dan Aritmia, meskipun masih terdapat kesalahan klasifikasi yang cukup signifikan, terutama adanya pertukaran prediksi antara kedua kelas tersebut. Sebaliknya, model tidak mampu mengklasifikasikan kelas Gagal Jantung secara tepat karena seluruh kasus pada kelas ini salah diprediksi, yang disebabkan oleh keterbatasan jumlah sampel. Pola kesalahan yang muncul mengindikasikan adanya kemiripan karakteristik klinis antar beberapa kelas penyakit jantung, khususnya antara Normal dan Aritmia, serta menegaskan bahwa ketidakseimbangan data dan kompleksitas gejala menjadi tantangan utama dalam klasifikasi penyakit kardiovaskular.

Hasil evaluasi yang terlihat pada Gambar 6 model *Local Mean K-Nearest Neighbor* dengan nilai K optimal sebesar 11 menunjukkan performa yang cukup baik dalam mengklasifikasikan penyakit jantung, dengan nilai *accuracy* sebesar 71,60%, *precision* 69,21%, *recall* 71,60%, dan *F1-score* 70,27%, yang mencerminkan keseimbangan yang relatif baik antara kemampuan deteksi dan kesalahan klasifikasi. Analisis per kelas menunjukkan bahwa model memiliki kinerja sangat baik pada kelas Penyakit Jantung Koroner dengan *precision* 0.92, *recall* 0.97, dan *F1-score* 0.94, sementara kelas Normal menunjukkan performa yang cukup stabil. Namun, performa menurun pada kelas Aritmia dan sangat terbatas pada kelas Gagal Jantung akibat jumlah sampel yang sangat sedikit. Perbedaan nilai *macro average* dan *weighted average* mengindikasikan adanya pengaruh *class imbalance* dalam dataset, sehingga meskipun model efektif pada kelas dengan representasi data yang memadai, diperlukan strategi tambahan seperti *oversampling* atau *cost-sensitive learning* untuk meningkatkan performa pada kelas minoritas.

4. PEMBAHASAN

Pada penelitian ini, algoritma *Local Mean K-Nearest Neighbor* (*LMKNN*) diterapkan untuk mengklasifikasikan penyakit jantung berdasarkan data rekam medis pasien yang memiliki karakteristik kompleks dan rentan terhadap *noise* serta *outlier*. Penerapan *LMKNN* terbukti mampu mengatasi kelemahan *K-Nearest Neighbor* (*KNN*) tradisional dengan memanfaatkan pendekatan *local mean*, sehingga representasi tetangga terdekat menjadi lebih stabil dan tidak terlalu dipengaruhi oleh data ekstrem. Proses *preprocessing* yang mencakup pembersihan data, *feature encoding*, dan *feature scaling* berperan penting dalam meningkatkan kualitas data sebelum proses klasifikasi, yang secara langsung berdampak pada performa model.

Hasil *hyperparameter tuning* menunjukkan bahwa nilai $K=11$ merupakan parameter optimal dengan nilai *cross-validation accuracy* sebesar 75,5%. Temuan ini mengindikasikan bahwa pemilihan jumlah tetangga yang tepat sangat berpengaruh

terhadap kinerja LMKNN, karena nilai K yang terlalu kecil dapat meningkatkan sensitivitas terhadap *noise*, sedangkan nilai K yang terlalu besar berpotensi mengaburkan karakteristik lokal data. Dengan nilai $K=11$, model mampu menyeimbangkan antara stabilitas dan sensitivitas terhadap pola data medis yang ada.

Evaluasi model pada data pengujian menghasilkan nilai *accuracy* sebesar 71,60%, *precision* 69,21%, *recall* 71,60%, dan *F1-score* 70,27%, yang menunjukkan bahwa model memiliki performa klasifikasi yang cukup baik secara keseluruhan. Secara khusus, model menunjukkan kinerja yang sangat unggul dalam mendeteksi kasus Penyakit Jantung Koroner dengan nilai *precision* 91,89% dan *recall* 97,14%. Tingginya nilai *recall* pada kelas ini mengindikasikan bahwa model sangat efektif dalam meminimalkan kesalahan *false negative*, yang sangat penting dalam konteks medis karena kesalahan tersebut dapat berdampak fatal terhadap pasien.

Analisis *feature importance* memperlihatkan bahwa variabel Nyeri Dada merupakan fitur paling dominan dalam proses klasifikasi dengan kontribusi sebesar 73,79%, diikuti oleh *Heart Rate (HR)* sebesar 36,40% dan *Respiratory Rate (RR)* sebesar 25,25%. Hasil ini selaras dengan kondisi klinis nyata, dimana nyeri dada merupakan gejala utama yang sering dijadikan indikator awal dalam diagnosis penyakit jantung. Dominasi fitur-fitur vital tersebut menunjukkan bahwa model berhasil menangkap hubungan yang relevan antara parameter klinis dan kondisi kardiovaskular pasien.

Meskipun demikian, penelitian ini juga mengungkap keterbatasan dalam menangani *class imbalance*, khususnya pada kelas Gagal Jantung yang memiliki jumlah sampel sangat terbatas. Kondisi ini menyebabkan nilai *precision*, *recall*, dan *F1-score* pada kelas tersebut mencapai 0,00%, yang menunjukkan bahwa model belum mampu mengenali pola kelas minoritas secara optimal. Hal ini menegaskan pentingnya penerapan strategi penanganan ketidakseimbangan data, seperti *oversampling*, *undersampling*, atau pendekatan berbasis *cost-sensitive learning*, pada penelitian lanjutan.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma *Local Mean K-Nearest Neighbor* memiliki potensi yang baik untuk digunakan dalam klasifikasi penyakit jantung berbasis data rekam medis. Dukungan implementasi sistem prediksi *real-time* dengan penyajian *confidence score* dan distribusi probabilitas semakin memperkuat peran model sebagai *clinical decision support tool*. Dengan pengembangan lebih lanjut, khususnya dalam penanganan ketidakseimbangan kelas dan perluasan dataset, model ini berpotensi memberikan kontribusi yang signifikan dalam mendukung proses diagnosis penyakit jantung secara lebih akurat dan andal.

5. KESIMPULAN

Algoritma *Local Mean K-Nearest Neighbor (LMKNN)* berhasil diterapkan secara efektif dalam mengklasifikasikan penyakit jantung berdasarkan data rekam medis dengan pendekatan *local mean* yang mampu mengurangi pengaruh *noise* dan *outliers*. Hasil *hyperparameter tuning* menunjukkan bahwa nilai $K=11$

memberikan performa terbaik dengan *cross-validation accuracy* sebesar 75,5%. Pada data pengujian, model mencapai kinerja yang cukup baik dengan nilai *accuracy* 71,60%, *precision* 69,21%, *recall* 71,60%, dan *F1-score* 70,27%, serta menunjukkan performa yang sangat unggul dalam mendeteksi Penyakit Jantung Koroner dengan *recall* 97,14% dan tingkat *false negative* yang sangat rendah. Pembagian data menggunakan *stratified sampling* 80:20 berhasil menjaga proporsi kelas, sementara analisis *feature importance* mengidentifikasi Nyeri Dada sebagai fitur paling berpengaruh, diikuti oleh HR dan RR. Meskipun demikian, model masih menghadapi keterbatasan dalam menangani *class imbalance* pada kelas minoritas Gagal Jantung, sehingga diperlukan strategi penanganan lebih lanjut. Secara keseluruhan, sistem prediksi yang dikembangkan memiliki potensi sebagai *clinical decision support tool* dalam membantu diagnosis penyakit jantung secara *real-time* berdasarkan parameter klinis.

DAFTAR PUSTAKA

- [1] A. Putra, D. Abdullah, and M. Daud, "Menggunakan Metode K-Harmonic Means dengan Validasi Silhouette Index dan C-index," *Jurnal Janitra Informatika dan Sistem Informasi*, vol. 4, no. 2, pp. 74–86, 2024, doi: [10.59395/janitra.v4i2.203](https://doi.org/10.59395/janitra.v4i2.203).
- [2] S. Asyuti and A. A. Setyawan, "Data Mining Dalam Penggunaan Presensi Karyawan Denga Cluster Means," *Jurnal Ilmiah Sains Teknologi Dan Informasi*, vol. 1, no. 1, pp. 01–10, 2023.
- [3] A. A. M. Lubis, R. K. Dinata, and H. A. K. Aidilof, "Classification of Heart Disease Using Modified K-Nearest Neighbor (MKNN) Method," *Journal of Advanced Computer Knowledge and Algorithms*, vol. 1, no. 2, p. 31, 2024, doi: [10.29103/jacka.v1i2.15702](https://doi.org/10.29103/jacka.v1i2.15702).
- [4] A. Yogiarto, A. Homaidi, and Z. Fatah, "Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung," *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 3, pp. 1720–1728, 2024, doi: [10.33379/gtech.v8i3.4495](https://doi.org/10.33379/gtech.v8i3.4495).
- [5] H. H. Sinaga and S. Agustian, "Pebandingan Metode Decision Tree dan XGBoost untuk Klasifikasi Sentimen Vaksin Covid-19 di Twitter," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 8, no. 3, pp. 107–114, Dec. 2022, doi: [10.25077/teknosi.v8i3.2022.107-114](https://doi.org/10.25077/teknosi.v8i3.2022.107-114).
- [6] T. Angelya, A. Rahman, and I. Pradesan, "Implementasi Sistem Klasifikasi Tim Kerja Menggunakan Algoritma Naïve Bayes (Studi Kasus: PT.MNL)," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 2, pp. 86–94, Aug. 2023, doi: [10.25077/teknosi.v9i2.2023.86-94](https://doi.org/10.25077/teknosi.v9i2.2023.86-94).
- [7] A. S. Hidayatullah, F. A. Bachtiar, and I. Cholissodin, "Penerapan Algoritme Nearest Centroid Neighbor Classifier Based on K Local Means Using Harmonic Mean Distance (LMKHNCN) Untuk Klasifikasi Hasil Kinerja Pegawai Negeri Sipil," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 6, pp. 1287–1292, 2021, doi: [10.25126/jtiik.2021834431](https://doi.org/10.25126/jtiik.2021834431).
- [8] R. Naifa Sanjy *et al.*, "Penerapan Algoritma K-Nearest Neighbor pada Klasifikasi Penyakit Jantung," *BIAStatistics Journal of Statistics Theory and*

- Application*, Jan. 2022, [Online]. Available: <http://prosiding.statistics.unpad.ac.id>
- [9] D. Cahya Putri Buani, "Penerapan Algoritma Naïve Bayes dengan Seleksi Fitur Algoritma Genetika Untuk Prediksi Gagal Jantung," *EVOLUSI: Jurnal Sains dan Manajemen*, vol. 9, no. 2, pp. 43–48, 2021, doi: [10.31294/evolusi.v9i2.11141](https://doi.org/10.31294/evolusi.v9i2.11141).
- [10] D. H. Depari, Y. Widiastiwi, and M. M. Santoni, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *Informatik: Jurnal Ilmu Komputer*, vol. 18, no. 3, p. 239, 2022, doi: [10.52958/iftk.v18i3.4694](https://doi.org/10.52958/iftk.v18i3.4694).
- [11] R. Riky, D. Destinady, F. R. W. P. F. Yani, and R. Jannah, "Penanganan Penyakit Jantung Pada Ibu Hamil," *Jurnal Forum Kesehatan: Media Publikasi Kesehatan Ilmiah*, vol. 11, no. 1, pp. 20–25, 2022, doi: [10.52263/jfk.v11i1.223](https://doi.org/10.52263/jfk.v11i1.223).
- [12] D. Pradana, M. Luthfi Alghifari, M. Farhan Juna, and D. Palaguna, "Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network," *Indonesian Journal of Data and Science*, vol. 3, no. 2, pp. 55–60, 2022, doi: [10.56705/ijodas.v3i2.35](https://doi.org/10.56705/ijodas.v3i2.35).
- [13] S. D. Lestari, I. S. Wahyuningsih, and A. I. Amal, "Gambaran aktivitas fisik dan tingkat kenyamanan pasien dengan penyakit jantung koroner," *Jurnal Ilmiah Sultan Agung*, vol. 2, no. 1, pp. 575–582, 2023.
- [14] S. Adi and A. Wintarti, "Komparasi Metode Support Vector Machine (Svm), K-Nearest Neighbors (Knn), Dan Random Forest (Rf) Untuk Prediksi Penyakit Gagal Jantung," *MATHunesa: Jurnal Ilmiah Matematika*, vol. 10, no. 2, pp. 258–268, 2022, doi: [10.26740/mathunesa.v10n2.p258-268](https://doi.org/10.26740/mathunesa.v10n2.p258-268).
- [15] Niendy Alexandra Yosephine and Ratnadewi, "Penggunaan Artificial Neural Network pada Sinyal Elektrokardiogram untuk Mendeteksi Penyakit Jantung Aritmia Supraventrikular," *INFORMASI (Jurnal Informatika dan Sistem Informasi)*, vol. 13, no. 1, pp. 14–23, 2021, doi: [10.37424/informasi.v13i1.69](https://doi.org/10.37424/informasi.v13i1.69).
- [16] A. Suryadi, Y. W. T. Arif, and N. S. Novitasari, "Rancang Bangun Sistem Informasi Rekam Medis Klinik Rawat Jalan Berbasis Web," *Infokes: Jurnal Ilmiah Rekam Medis dan Informatika Kesehatan*, vol. 12, no. 1, pp. 37–43, 2022, doi: [10.47701/infokes.v12i1.1498](https://doi.org/10.47701/infokes.v12i1.1498).
- [17] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, "Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner," *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 1, Apr. 2021, doi: [10.30813/jbase.v4i1.2729](https://doi.org/10.30813/jbase.v4i1.2729).
- [18] A. D. Alfarizi and A. Andri, "Pemanfaatan Data Mining Dalam Memprediksi Produksi Pada PT Pupuk Sriwidjaja Palembang Menggunakan Algoritma Regresi Linier Berganda," *Jurnal Nasional Ilmu Komputer*, vol. 2, no. 1, pp. 51–63, 2021, doi: [10.47747/jurnalnik.v2i1.522](https://doi.org/10.47747/jurnalnik.v2i1.522).
- [19] H. Nisa, M. Daud, and S. Retno, "JOISIE licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0) Evaluasi Kinerja K-Medoids Clustering Model Untuk Klasterisasi Daerah Produktivitas Panen Padi Di Kabupaten Bireuen," *Journal Of Information Systems And Informatics Engineering*, vol. 9, no. 2, 2025, doi: [10.35145/joisie.v9i2.4965](https://doi.org/10.35145/joisie.v9i2.4965).
- [20] M. F. Haryanti *et al.*, "Pengaruh Data Mining, Strategi Perusahaan Terhadap Laporan Kinerja Perusahaan," *Jurnal Manajemen dan Bisnis*, vol. 3, no. 1, pp. 71–90, 2024.
- [21] N. Putri Setyadini, "Penerapan Data Mining Untuk Prediksi Hasil Produksi Karet Menggunakan Algoritma Decision Tree C4.5," *Informatika*, vol. 2, no. 7, pp. 1–11, 2022.
- [22] Andi Diah Kuswanto, Said Imam Puro, Jodi Hariyan, Ridho Rafliansyah, Muhammad Rival Aziz, and Pebro Vaulina Rajagukguk, "Analisa Data Shopping Trends Menggunakan Algoritma Klasifikasi Dengan Metode Naive Bayes," *Repeater: Publikasi Teknik Informatika dan Jaringan*, vol. 2, no. 3, pp. 119–134, 2024, doi: [10.62951/repeater.v2i3.118](https://doi.org/10.62951/repeater.v2i3.118).
- [23] A. H. Nasrullah, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik," *Jurnal Pilar Nusa Mandiri*, vol. 7, no. 2, p. 217, 2021.
- [24] Y. Fatimah, Asrianda, and R. Meiyanti, "Klasifikasi Jenis Tanah Yang Sesuai Terhadap Tanaman Pangan Menggunakan Metode Random Forest Di Kabupaten Mandailing Natal," *Seminar Nasional Teknologi Informasi dan Teknik Informatika (SENASTIKA)*, vol. 1, no. 1, pp. 1–9, 2024.