

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data mining merupakan suatu proses sistematis yang berfokus pada eksplorasi dan analisis data dengan tujuan mengekstraksi atau menyaring informasi dari kumpulan data berskala besar. Proses ini dilakukan melalui serangkaian tahapan yang terstruktur guna memperoleh wawasan yang bernilai (Satriono *et al.*, 2025). 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 (Lubis *et al.*, 2024).

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 oleh rusaknya sel otot dalam jantung yang memompa dan membawa aliran darah keseluruh tubuh. Kerusakan tersebut diawali dengan adanya proses *ateroskelosis* yang merupakan proses kompleks yang mengakibatkan terjadinya pengendapan terhadap proliferasi elemen yang terdapat pada dinding arteri serta terjadinya *lipoprotein plasma*. Kondisi tersebut terus mengalami perkembangan dengan rangkaian tahap yang diawali dari lemak yang terdapat sel busa hingga secara terus-menerus berkembang hingga akhirnya terdapat timbunan plak dan terdapat lesi jaringan ikat yang menutup. plak tersebut dapat menjadi penghalang dari mengalirnya darah arteri sehingga dapat memicu terjadinya peristiwa klinis (Yogianto *et al.*, 2024).

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. Teknik ini digunakan untuk mengkategorikan objek ke dalam kelompok tertentu berdasarkan perilaku dan atribut yang telah didefinisikan sebelumnya. 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 memperhitungkan karakteristik lokal setiap kelas secara lebih spesifik (Rosdiana, 2024).

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* (Akbarollah *et al.*, 2023), 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).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, permasalahan dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana penerapan metode *Local Mean K-Nearest Neighbor* (LMKNN) dalam klasifikasi penyakit jantung?
2. Sejauh mana tingkat akurasi yang dapat dicapai oleh metode *Local Mean K-Nearest Neighbor* (LMKNN) dalam mengklasifikasikan penyakit jantung?

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Menerapkan algoritma *Local Mean K-Nearest Neighbor* (LMKNN) dalam proses klasifikasi penyakit jantung.
2. Menguji dan menganalisis tingkat akurasi *Local Mean K-Nearest Neighbor* dalam mengklasifikasikan penyakit jantung.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Berkontribusi dalam pengembangan metode klasifikasi penyakit jantung yang lebih akurat dengan menerapkan algoritma *Local Mean K-Nearest Neighbor* (LMKNN), sehingga dapat digunakan sebagai dukungan dalam proses diagnosis berbasis data rekam medis.
2. Meningkatkan akurasi dalam proses diagnosis penyakit jantung berdasarkan analisis data rekam medis, sehingga dapat menjadi alat pendukung yang efektif dalam pengambilan keputusan medis yang lebih akurat dan terpercaya.
3. Memberikan dasar bagi penelitian lanjutan dalam pengembangan algoritma klasifikasi penyakit menggunakan data rekam medis dengan pendekatan yang lebih efektif.
4. Memudahkan pasien dalam mengetahui diagnosa awal secara jelas.

1.5 Ruang Lingkup dan Batasan Masalah

Lingkup permasalahan dalam penelitian ini ditetapkan sebagai berikut:

1. Penelitian ini terbatas pada klasifikasi penyakit jantung dengan menggunakan data rekam medis yang tersedia sebagai sumber informasi.
2. Algoritma yang diterapkan dalam penelitian ini adalah *Local Mean K-Nearest Neighbor* (LMKNN), yang merupakan pengembangan dari algoritma KNN, tanpa membahas implementasi metode klasifikasi lainnya untuk penyakit jantung.
3. Penelitian ini tidak mencakup implementasi dalam sistem nyata, tetapi hanya berfokus pada pengembangan model klasifikasi dan evaluasi algoritma LMKNN dalam konteks penyakit jantung.

4. Fitur yang diteliti mencakup delapan jenis yaitu jenis kelamin, usia, tekanan darah, *heart rate* (HR), *respiratory rate* (RR), hasil elektrokardiografi (EKG), klasifikasi penyakit, serta tingkat nyeri dada.
5. Label yang dihasilkan dari penelitian ini ada empat yaitu jantung koroner, gagal jantung, aritmia, dan normal.
6. Evaluasi model klasifikasi dilakukan menggunakan confusion matrix dengan parameter *accuracy*, *recall*, F-1 score.