

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Penyakit jantung merupakan kondisi medis yang kompleks dan serius yang berdampak langsung pada jantung serta sistem kardiovaskular secara keseluruhan. Gangguan pada arteri koroner dan komplikasi lain sering menjadi penyebab utama terganggunya fungsi jantung, yang dapat berujung pada morbiditas dan mortalitas tinggi di seluruh dunia. Oleh karena itu, diagnosis dini yang akurat sangat krusial untuk mencegah komplikasi yang lebih serius dan meningkatkan prognosis pasien [1]. Menurut data dari Organisasi Kesehatan Dunia (WHO), penyakit jantung menyumbang lebih dari 30% dari total kematian global setiap tahunnya [2]. Di Indonesia sendiri, prevalensi penyakit ini juga menunjukkan tren peningkatan yang signifikan, sehingga upaya deteksi dini dengan metode yang efektif menjadi sangat penting untuk mengurangi dampak negatif bagi kualitas hidup pasien [3].

Kemajuan terbaru dalam teknologi kecerdasan buatan (Artificial Intelligence/AI), khususnya penerapan machine learning (ML), membuka peluang besar dalam bidang diagnostik medis. Machine learning memungkinkan pemodelan prediktif yang dapat memproses data rekam medis yang besar dan kompleks secara efisien untuk mendeteksi tanda-tanda penyakit jantung sejak tahap awal [4] [5]. Di antara beragam algoritma machine learning, K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) merupakan dua metode yang paling banyak digunakan dan diteliti untuk klasifikasi penyakit jantung. Algoritma KNN bekerja dengan mengklasifikasikan data berdasarkan kedekatan atau jarak terhadap k tetangga terdekat dalam ruang fitur, sehingga cocok untuk dataset yang distribusinya jelas, meskipun performanya bisa menurun pada data berdimensi tinggi atau besar. Beberapa varian seperti Cubic KNN menerapkan transformasi non-linear untuk mempertahankan jarak lokal sekaligus meningkatkan akurasi klasifikasi. Di sisi lain, SVM mencari hyperplane optimal yang memaksimalkan margin antar kelas, membuatnya sangat efektif untuk mengklasifikasikan data berdimensi tinggi dan kompleks, termasuk masalah klasifikasi linear maupun non-linear, dengan stabilitas hasil dan kemampuan generalisasi yang baik [6] [7].

Studi oleh Anggoro, D. A et al (2020) menegaskan pentingnya proses preprocessing, khususnya normalisasi data, dalam meningkatkan performa model klasifikasi, terutama pada dataset berdimensi tinggi dan kompleks [8]. Dalam aplikasi nyata, penelitian terbaru oleh Sevkett Ay et al. (2023) melaporkan optimasi fitur berbasis meta-heuristik pada KNN yang mampu mencapai F-score sebesar 99,72% pada dataset Cleveland, meskipun terdapat keterbatasan dalam menangani data berdimensi sangat tinggi [9]. Sebaliknya, studi oleh C.M.M. Mansoor et al. (2024) menggarisbawahi keunggulan SVM dalam memproses data berdimensi tinggi dan non-linear dengan efisiensi waktu yang lebih baik serta sensitivitas tinggi terhadap klasifikasi penyakit jantung [10]. Perbandingan ini menggambarkan bahwa meskipun SVM secara umum lebih stabil dan memiliki kapabilitas generalisasi yang kuat, KNN tetap relevan sebagai metode yang efektif selama parameter-parameter model dioptimalkan secara tepat.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan melakukan **analisa komparasi dengan algoritma K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) untuk prediksi penyakit jantung**. Dengan membandingkan kedua algoritma tersebut, penelitian berharap dapat menentukan metode yang paling akurat dan efisien sehingga memberikan kontribusi signifikan bagi peningkatan diagnosis dini penyakit jantung.

1.2. Rumusan Masalah

Permasalahan yang akan dibahas dalam penelitian ini adalah bagaimana membandingkan kinerja algoritma K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) dalam melakukan prediksi penyakit jantung berdasarkan data pasien.

1.3. Tujuan dan Manfaat

1.3.1. Tujuan

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) dalam melakukan prediksi penyakit jantung berdasarkan data pasien, sehingga dapat diketahui algoritma mana yang paling akurat, efektif, dan tepat digunakan dalam prediksi penyakit jantung.

1.3.2. Manfaat

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut :

1. Memberikan kontribusi dalam pengembangan metode klasifikasi penyakit jantung berbasis machine learning dengan melakukan evaluasi komparatif antara algoritma K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM), sehingga menghasilkan rekomendasi algoritma yang paling efektif untuk diterapkan di bidang medis.
2. Mendukung tenaga medis dalam pengambilan keputusan klinis, khususnya dalam prediksi dan diagnosis penyakit jantung, melalui penerapan teknologi klasifikasi yang akurat dan efisien, sehingga dapat meningkatkan kualitas pelayanan kesehatan.
3. Menjadi acuan bagi penelitian selanjutnya dalam bidang pengolahan data medis dan pengembangan algoritma prediksi penyakit, dengan menyediakan analisa mendalam mengenai kelebihan dan kekurangan algoritma KNN dan SVM pada kasus prediksi penyakit jantung.

1.4. Batasan Masalah

Penelitian ini akan terfokus dalam mengklasifikasikan tingkat kecemasan dengan batasan-batasan masalah sebagai berikut:

1. Penelitian ini hanya menggunakan dataset penyakit jantung yang diperoleh dari situs Kaggle dengan link [kaggle.com](https://www.kaggle.com) sebagai sumber data utama.
2. Fokus penelitian dilekatkan pada analisis perbandingan kinerja dua algoritma machine learning, yaitu K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM), dalam memprediksi penyakit jantung berdasarkan fitur-fitur yang tersedia pada dataset.
3. Penelitian ini membatasi evaluasi algoritma berdasarkan metrik klasifikasi seperti akurasi, presisi, recall, dan F1-score tanpa memperluas ke aspek lain seperti optimasi parameter lanjutan atau penggunaan algoritma lain di luar KNN dan SVM.
4. Penelitian ini tidak mengkaji faktor eksternal yang dapat mempengaruhi prediksi penyakit jantung di dunia nyata, seperti kondisi medis pasien yang tidak tercakup dalam dataset atau variasi data yang lebih luas di luar dataset yang digunakan.

1.5. Keterbaruan

1. Penelitian yang dilakukan oleh Zhang et al. (2021) memperkenalkan metode prediksi penyakit jantung dengan menggunakan algoritma Support Vector Machine (SVM) dikombinasikan teknik seleksi fitur untuk meningkatkan akurasi prediksi. Hasil penelitian menunjukkan peningkatan akurasi klasifikasi hingga mencapai 85%, menekankan pentingnya pemilihan fitur pada data medis untuk hasil yang lebih valid dan akurat [11].
2. Penelitian yang dilakukan oleh E. M. A. Allah et al. (2022) mengusulkan optimasi parameter berbasis metaheuristik pada algoritma K-Nearest Neighbor (KNN), berhasil meningkatkan akurasi prediksi penyakit jantung hingga 86,8%. Studi ini pula membandingkan KNN dengan Support Vector Machine (SVM) yang mencapai akurasi 90,9%, menunjukkan keunggulan SVM dalam menangani data tidak seimbang [12].
3. Penelitian oleh Kanak Kalita et al. (2023) mengusulkan teknik Hybrid Deep Belief Network dan XGBoost (HDBN-XG) yang berhasil meningkatkan akurasi prediksi penyakit jantung hingga 99% dengan presisi 95%, menggunakan metode pemilihan fitur Multi-Objective Artificial Bee Colony. Pendekatan ini menonjolkan kemampuan mengatasi data tidak seimbang dan meningkatkan sensitivitas serta spesifisitas model prediksi, serta menekankan pentingnya fitur relevan yang dipilih melalui optimasi cerdas [13].
4. Penelitian oleh Sharma et al. (2024) mengembangkan model prediksi penyakit jantung berbasis ensemble machine learning yang menggabungkan beberapa algoritma, dengan teknik preprocessing data matang untuk menangani data tidak seimbang. Model Random Forest dan K-Nearest Neighbor mencapai akurasi tertinggi sebesar 84,15%, sedangkan Logistic Regression dan Support Vector Machine masing-masing mencapai akurasi 83,81% dan 82,49% [14].
5. Penelitian oleh Gufran Ahmad Ansari et al. (2023) mengevaluasi enam teknik machine learning untuk prediksi penyakit jantung, menunjukkan bahwa Random Forest dan K-Nearest Neighbor memberikan akurasi tertinggi hingga 99,04%. Studi ini juga menyoroti perlunya fitur seperti usia, tekanan darah, kolesterol, serta status perokok dalam peningkatan akurasi model [15].