

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Penyakit jantung adalah salah satu penyebab utama kematian di seluruh dunia, dengan laporan Organisasi Kesehatan Dunia (WHO) menyebutkan bahwa pada tahun 2017 sekitar 17,90 juta kematian terjadi akibat penyakit jantung [1]. Deteksi dini yang cepat dan akurat sangat penting untuk mengurangi angka kematian ini. Namun, metode tradisional seperti Elektrokardiogram (EKG) masih memerlukan waktu dan keterlibatan manual dari ahli medis, sehingga belum efisien dalam menangani deteksi dini penyakit jantung [2].

Perkembangan ilmu komputer, khususnya dalam bidang *machine learning*, telah membuka peluang baru dalam mendiagnosis penyakit jantung secara otomatis. Algoritma seperti *Random Forest* dan *Gradient Boosting Machine* sudah banyak diterapkan untuk meramalkan risiko penyakit jantung dengan memeriksa data pasien yang meliputi informasi demografis dan klinis[3]. Penggunaan *machine learning* membantu proses diagnosis menjadi lebih cepat dan lebih akurat daripada metode manual [4].

Namun, meskipun penggunaan *machine learning* telah berkembang pesat, masih ada tantangan dalam meningkatkan akurasi dan efisiensi deteksi dini penyakit jantung [5]. Oleh karena itu, penelitian ini bertujuan membandingkan kinerja dua algoritma yang sudah dikenal, yaitu *Random Forest* dan *Gradient Boosting Machine*, dalam melakukan prediksi terhadap faktor risiko penyakit jantung.

Teknologi yang digunakan saat ini untuk diagnosis penyakit jantung mencakup *Random Forest* dan *Gradient Boosting Machine*. Untuk meningkatkan akurasi prediksi, *Random Forest* membangun beberapa *decision trees* dan menggabungkannya, sementara *Gradient Boosting Machine* menggunakan pendekatan iteratif untuk mengoptimalkan hasil prediksi dengan memperbaiki kesalahan secara bertahap [6][7]. Kedua algoritma ini mampu mengolah data medis pasien seperti usia, jenis kelamin, kadar gula darah, dan riwayat kesehatan [8].

Sistem pendukung keputusan berbasis *machine learning* telah membantu dokter dalam mempercepat proses diagnosis dan memberikan hasil prediksi yang lebih

akurat. Meskipun begitu, teknologi ini belum mampu sepenuhnya menggantikan keterlibatan ahli medis, terutama dalam kasus yang lebih kompleks [9].

Teknologi terbaru yang sedang dikembangkan adalah model *deep learning*, khususnya *Convolutional Neural Networks* (CNNs), yang dapat mengolah data medis lebih kompleks seperti citra medis (contohnya, MRI dan CT scan). CNNs memungkinkan analisis yang lebih dalam dan lebih akurat dalam mendeteksi pola yang sulit dikenali oleh manusia atau algoritma *machine learning* tradisional [9]. Namun, teknologi ini belum banyak diterapkan secara luas karena tantangan besar dalam hal kebutuhan komputasi yang tinggi, jumlah data yang besar, dan kesulitan dalam interpretasi hasil prediksi [10].

Penelitian ini mengusulkan penggunaan algoritma *Random Forest* dan *Gradient Boosting Machine* untuk memprediksi faktor risiko penyakit jantung. Kedua algoritma ini dipilih karena kemudahan penerapannya dan efisiensi komputasi yang lebih tinggi dibandingkan teknologi *deep learning*. *Random Forest* efektif dalam menangani data dengan banyak variabel independen, sedangkan *Gradient Boosting Machine* bekerja secara iteratif untuk mengoptimalkan hasil prediksi [11].

Algoritma ini cocok digunakan di rumah sakit atau klinik dengan keterbatasan infrastruktur komputasi, namun tetap memerlukan hasil diagnosis yang cepat dan akurat. Dengan metode ini, dokter dapat dengan mudah memahami dan menggunakan hasil prediksi untuk pengambilan keputusan klinis [12]

Teknologi *deep learning* memerlukan data yang lebih besar dan komputasi yang lebih tinggi, tetapi sulit untuk diinterpretasikan oleh pengguna non-teknis [10]. Sebaliknya, model *Random Forest* dan *Gradient Boosting Machine* lebih mudah dipahami, lebih cepat diterapkan, dan dapat memberikan hasil yang kompetitif dengan biaya komputasi yang lebih rendah [13]. Kedua algoritma ini lebih sesuai untuk analisis data medis tabular dan memberikan hasil yang dapat diinterpretasikan dengan jelas oleh dokter, sehingga membantu mempercepat dan mempermudah proses pengambilan keputusan klinis [14].

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana cara mengidentifikasi dan mendiagnosis penyakit jantung kardiovaskular pada pasien menggunakan atribut medis seperti jenis kelamin, usia, nyeri dada, dan kadar gula puasa?
2. Apa saja faktor risiko yang berkontribusi terhadap penyakit jantung, baik yang dapat diubah maupun yang tidak dapat diubah?

1.3. Tujuan dan Manfaat

1.3.1. Tujuan

Tujuan dari penelitian ini adalah untuk:

1. Mengidentifikasi dan mendiagnosis penyakit jantung kardiovaskular pada pasien dengan menggunakan atribut medis seperti jenis kelamin, usia, nyeri dada, dan kadar gula puasa.
2. Menganalisis faktor risiko yang berkontribusi terhadap penyakit jantung, baik yang dapat diubah maupun tidak dapat diubah.

1.3.2. Manfaat

Manfaat dari penelitian ini meliputi:

1. Memberikan wawasan yang lebih baik tentang faktor-faktor yang mempengaruhi kesehatan jantung, sehingga dapat membantu dalam pencegahan dan pengelolaan penyakit jantung.
2. Meningkatkan akurasi diagnosis penyakit jantung melalui penerapan algoritma pembelajaran mesin, yang dapat mengurangi waktu dan upaya yang diperlukan dalam proses diagnosis.

1.4. Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya menggunakan dataset tertentu yang mungkin tidak mencakup seluruh populasi pasien dengan penyakit jantung.
2. Penelitian ini hanya mempertimbangkan atribut medis tertentu seperti jenis kelamin, usia, nyeri dada, dan kadar gula puasa.
3. Penelitian ini terbatas pada penggunaan algoritma *Random Forest* dan *Gradient Boosting Machine*.

4. Pengelompokan data yang digunakan sebagai strategi klasifikasi mungkin tidak mencakup semua metode yang ada,
5. Dataset yang digunakan berasal dari
<https://www.kaggle.com/datasets/mexwell/heart-disease-dataset/data>

1.5. Keterbaruan

1. Menurut (Omotehinwa et al., 2024), Dalam penelitian ini menawarkan pendekatan inovatif dalam menganalisis dan membandingkan model *Random Forest* dan *Gradient Boosting Machine* untuk meningkatkan akurasi prediksi faktor risiko Penyakit Jantung Koroner (PJK), dengan memanfaatkan teknik pengambilan sampel *Borderline-SMOTE* dan optimasi hiperparameter menggunakan metode Bayesian, yang dapat berkontribusi pada pengembangan sistem deteksi dini yang lebih efektif untuk PJK.
2. Menurut (Asadi et al., 2021), Dalam penelitian ini terletak pada penerapan teknik data mining yang inovatif untuk meningkatkan akurasi diagnosis penyakit jantung melalui metode *ensemble* yang menggabungkan berbagai classifier, sehingga memberikan pendekatan yang lebih efektif dalam memprediksi penyakit kardiovaskular.
3. Menurut (Ramesh et al., 2022), Dalam penelitian ini terletak pada pengembangan metode inovatif berbasis *machine learning* untuk memprediksi penyakit jantung, yang mengintegrasikan berbagai teknik pemilihan fitur dan validasi silang untuk meningkatkan akurasi dan efisiensi klasifikasi.
4. Menurut (Ogunpola et al., 2024), Dalam penelitian ini terletak pada penerapan berbagai teknik *machine learning* untuk meningkatkan akurasi deteksi penyakit kardiovaskular, yang merupakan tantangan signifikan dalam bidang kesehatan saat ini
5. Menurut (Çakmak, 2024), penelitian ini terletak pada penerapan algoritma *machine learning Random Forest* yang menunjukkan akurasi 99% dalam klasifikasi penyakit jantung, mengungguli metode lain dan berpotensi meningkatkan diagnosis serta hasil pasien dalam pengobatan kardiovaskular.