

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Penyakit jantung (*cardiovaskular*) merupakan salah satu penyakit kronis yang sangat berbahaya [1]. Jika tidak ditangani dengan baik dan benar, penyakit ini dapat menjadi penyebab kematian utama di dunia [2]. Menurut data dari *World Health Organization* (WHO), pada tahun 2021 sekitar 39 juta atau 57% dari total 68 juta kematian yang ada di seluruh dunia disebabkan oleh 10 penyakit utama. Dua kelompok paling besar penyebab kematian terbanyak berasal dari penyakit jantung dan pembuluh darah (jantung iskemik dan stroke), serta penyakit pernapasan. Sebanyak 13% dari total kematian disebabkan oleh penyakit tidak menular seperti penyakit jantung [3].

Fokus utama, untuk pencegahan dan pengendalian penyakit jantung adalah dengan menjaga kesehatan masyarakat serta mencegah munculnya faktor yang dapat berisiko [4]. Pada penyakit jantung, faktor risiko dapat diklasifikasikan menjadi 2 yaitu [5], yang tidak dimodifikasi, seperti riwayat keluarga, dan jenis kelamin. Dan yang dapat dimodifikasi, seperti hipertensi, diabetes mellitus, merokok, dislipidemia, pola gaya hidup yang tidak sehat, dan aktifitas fisik yang rendah [6].

Kenyataan di lapangan, secara global penyakit *cardiovascular* masih menjadi masalah kesehatan [7] dan menyebabkan sekitar 9,4 juta kematian setiap tahunnya, dan di Indonesia sendiri, sekitar 1,5% penduduk dari semua kelompok umur yang menderita penyakit jantung, yang artinya setiap 100 orang, setidaknya ada satu atau dua orang yang mengalami penyakit jantung [8]. Kondisi ini diperparah dengan kurangnya kesadaran masyarakat tentang pentingnya deteksi dini penyakit jantung, sehingga banyak kasus yang baru terdiagnosis ketika sudah dalam stadium lanjut [9].

Penelitian terdahulu, telah menunjukkan bahwa dengan menggunakan *machine learning* khususnya algoritma *Random Forest* memiliki potensi besar dan akurasi lebih baik dalam memprediksi penyakit jantung melalui optimalisasi parameter dan pemilihan fitur [10]. Studi lain, telah melakukan pengujian dengan tiga klasifikasi model yaitu: *Decision Tree*, *Naïve Bayes*, Dan *Random Forest*. Bahwa *Random Forest* merupakan model terbaik dalam pengklasifikasian penyakit jantung [11]. Penelitian terbaru, juga dilakukan dengan membandingkan *K-Nearest Neighbors (KNN)*, Dan

Random Forest. *Random Forest* lebih efektif dalam memberikan rekomendasi gaya hidup sehat guna mencegah penyakit jantung [12].

Namun demikian, terdapat gap penelitian yang signifikan dalam literatur yang ada. Pertama, sebagian besar penelitian masih fokus pada data klinis seperti usia, kolestrol, dan tekanan darah. Namun, belum secara eksplisit memberikan faktor gaya hidup seperti merokok, pola makan, obesitas, dan aktivitas fisik [13] Kedua, dengan menggunakan *Information Gain* sebagai seleksi fitur, dan belum menekankan interpretasi *Feature Importance* dalam memberikan konteks faktor berisiko [14]. Ketiga, belum memberikan secara khusus fokus terhadap faktor gaya hidup sebagai penentu dalam prediksi penyakit jantung [15].

Penelitian ini bertujuan untuk menganalisis faktor-faktor gaya hidup yang berpengaruh terhadap risiko penyakit jantung menggunakan algoritma *Random Forest* dengan *Feature Importance*. Pendekatan ini akan mengidentifikasi faktor gaya hidup yang memiliki kontribusi terbesar terhadap risiko penyakit jantung secara komprehensif dan terintegrasi. Melalui analisis *Feature Importance*, penelitian ini berfokus pada identifikasi faktor dominan seperti kebiasaan merokok, aktivitas fisik, dan pola makan. Dengan demikian, penelitian dengan judul “Analisis Faktor Gaya Hidup Terhadap Risiko Penyakit Jantung Menggunakan Random Forest Dengan Feature Importance” dapat memberikan kontribusi signifikan dalam meningkatkan akurasi prediksi penyakit jantung berbasis gaya hidup pada data pasien.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan penyakit jantung masih menjadi salah satu penyebab utama kematian di dunia, dengan tingginya angka kematian yang menunjukkan bahwa upaya pencegahan belum berjalan secara optimal. Salah satu penyebabnya bisa terjadi karena belum adanya pemahaman yang mendalam mengenai bagaimana faktor gaya hidup di masyarakat dapat memberikan dampak signifikan dalam menurunkan risiko penyakit jantung. Pola hidup seperti memperhatikan aktivitas fisik, dan kebiasaan sehari-hari sangat berpengaruh terhadap kesehatan jantung.

Oleh karena itu, penelitian ini dilakukan untuk memahami secara spesifik bagaimana hubungan faktor gaya hidup dengan risiko penyakit jantung. Selain itu, penelitian ini juga berfokus pada pembangunan model dengan Algoritma *Random Forest* dengan *Feature Importance* dengan optimal untuk menganalisis dan memprediksi faktor gaya hidup dalam prediksi risiko penyakit jantung. Dan dibutuhkan pendekatan inovatif dan integratif yang menggabungkan kekuatan teknologi *machine learning* serta data yang diperoleh dan diolah dari sumber *Kaggle* yang dapat diakses dan digunakan secara efektif dan baik.

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk menganalisis dan mengidentifikasi faktor-faktor gaya hidup yang berpengaruh terhadap risiko penyakit jantung menggunakan algoritma *Random Forest* dengan *Feature Importance*. Melalui pendekatan ini, penelitian ini berfokus pada identifikasi tingkat kepentingan faktor gaya hidup seperti kebiasaan merokok, aktivitas fisik, pola makan, dan faktor lainnya. Penelitian ini juga bertujuan untuk memberikan pemahaman yang lebih mendalam mengenai hubungan sebab-akibat antara gaya hidup dan risiko penyakit jantung untuk mendukung upaya pencegahan

1.4 Manfaat Penelitian

Beberapa manfaat yang diharapkan dari penelitian ini adalah:

1. Memberikan kontribusi signifikan pada pengembangan ilmu pengetahuan di bidang *machine learning* untuk kesehatan, khususnya dalam prediksi risiko penyakit *cardiovaskular* berbasis gaya hidup
2. Memperkaya literatur ilmiah mengenai penerapan algoritma *Random Forest* dengan *Feature Importance* dalam prediksi penyakit berbasis faktor perilaku dan gaya hidup
3. Mengembangkan pendekatan metodologi penelitian yang inovatif untuk penelitian serupa dengan fokus pada faktor gaya hidup
4. Memberikan panduan praktis dan komprehensif dalam *pre-processing*, *feature engineering*, dan analisis data gaya hidup untuk prediksi penyakit *cardiovaskular*

1.5 Batasan Masalah

Ruang lingkup dalam penelitian ini ialah:

1. Penelitian ini dibatasi pada analisis faktor-faktor gaya hidup dan tidak mencakup faktor genetik, riwayat medis keluarga yang kompleks, atau kondisi komorbiditas yang berat
2. Dataset yang digunakan terbatas pada data sekunder yang tersedia secara publik dan dapat diakses untuk keperluan penelitian akademik
3. Model yang dikembangkan menggunakan algoritma *Random Forest* sebagai metode utama dan tidak melakukan perbandingan dengan algoritma *machine learning* lain
4. Analisis *Feature Importance* dibatasi pada teknik yang tersedia dalam *Random Forest* dan tidak menggunakan semua metode *feature selection* yang komprehensif
5. Penelitian ini tidak mencakup implementasi sistem dalam bentuk aplikasi website atau aplikasi mobile yang siap digunakan oleh user

1.6 Keterbaruan

Keterbaruan dalam penelitian ini berdasarkan hasil kajian pustaka dari penelitian yang sebelumnya, diantaranya:

1. Penelitian yang dilakukan oleh Hidayat bersama dengan satu tim (2023) dengan judul “Klasifikasi Penyakit Jantung Menggunakan Random Forest Classifier” menunjukkan hasil studi lebih baik dengan menggunakan *Random Forest* yang mencapai 94% dibantu dengan teknik *pre-processing*, dan normalisasi data yang dapat menjadi alternatif dalam melakukan klasifikasi [16].
2. Berdasarkan penelitian dari Nur Halizah Alfajr, dan Sofi Defiyanti (2024) yang berjudul “Prediksi Penyakit Jantung Menggunakan Metode Random Forest Dan Penerapan Principal Component Analysis (Pca)”, meneliti dengan menggunakan data klinis dan reduksi dimensi dengan menggunakan Random Forest dan Principal Component Analysis (PCA) untuk meningkatkan akurasi prediksi penyakit jantung. Hasil menunjukkan akurasi yang sangat tinggi mencapai 98%, presisi 100%, F-1score 98%, dan recall 96% dalam memprediksi penyakit jantung [17].

3. Menurut jurnal Seftyan Wiguna bersama dengan satu tim (2025) dengan judul “Prediksi Harga Smartphone Berdasarkan Fitur Smartphone Dengan Random Forest Regression”, bahwa menggunakan algoritma *Random Forest* dalam memprediksi harga smathphone, menunjukkan kekuatan *Random Forest* dalam menangani data kompleks dan non-linear [18].
4. Berdasarkan penelitian Dita Setiawan bersama dengan teman satu tim (2025), menunjukkan *Random Forest* memiliki performa terbaik dengan akurasi 87,2%, dengan indikator utama dalam penelitian berupa *chest pain* (nyeri dada), kolestrol, dan tekanan darah [19].
5. Penelitian yang dilakukan oleh Akhmad Pandhu Wijaya (2025) dengan judul “Perbandingan Algoritma Klasifikasi Random Foresst dengan Naïve Bayes Classifier pada Studi Penyakit Berdasarkan Pola Nutrisi”, meneliti klasifikasi penyakit dengan membandingkan *Random Forest* dan *Naïve Bayes* berdasarkan pola nutrisi seperti kalori, gula, dan protein, yang lebih dominan dalam aspek nutrisi tanpa memperluas ke faktor lain seperti faktor gaya hidup. Menunjukkan hasil akurasi 92%, precision 91% F-1 score 90,5%, dan recall 90% lebih unggul dengan menggunakan algoritma *Random Forest* [20].