

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit jantung merupakan salah satu penyebab utama kematian di seluruh dunia. Menurut laporan dari World Health Organization (WHO), penyakit kardiovaskular bertanggung jawab atas sekitar 17,9 juta kematian setiap tahunnya, atau sekitar 31% dari total kematian global (World Health Organization, 2021). Di Indonesia, prevalensi penyakit jantung terus mengalami peningkatan dari tahun ke tahun. Berdasarkan data Riset Kesehatan Dasar (Riskesdas) yang diterbitkan oleh Kementerian Kesehatan Republik Indonesia, prevalensi penyakit jantung di Indonesia pada tahun 2013 tercatat sebesar 1,5% dan meningkat menjadi 1,9% pada tahun 2018. (Kementerian Kesehatan Republik Indonesia, 2019). Kondisi ini menunjukkan bahwa penyakit jantung masih menjadi ancaman serius terhadap kesehatan masyarakat, sehingga diperlukan upaya pencegahan dan deteksi dini yang lebih efektif.

Penyakit jantung dapat menyerang siapa saja tanpa memandang usia, jenis kelamin, maupun latar belakang sosial ekonomi. Faktor risiko penyakit ini secara umum dapat diklasifikasikan menjadi dua, yaitu faktor yang tidak dapat dimodifikasi seperti usia, jenis kelamin, dan faktor genetik, serta faktor yang dapat dimodifikasi seperti dislipidemia, hipertensi, diabetes melitus, kebiasaan merokok, kurangnya aktivitas fisik, dan pola makan yang tidak sehat (Husada et al., 2024). Dengan demikian, pengendalian terhadap faktor risiko yang dapat dimodifikasi menjadi sangat penting dalam upaya menekan angka kematian akibat penyakit jantung.

Deteksi dini memiliki peran yang sangat krusial dalam mengidentifikasi faktor risiko maupun gejala awal penyakit jantung sebelum berkembang menjadi kondisi yang lebih berat. Langkah ini dapat mencegah komplikasi lanjutan dan meningkatkan efektivitas intervensi medis yang dilakukan (Rumah Sakit PELNI, 2022). Seiring dengan perkembangan teknologi di bidang kesehatan, pemanfaatan metode analisis data dan kecerdasan buatan mulai banyak diterapkan untuk meningkatkan akurasi dalam deteksi dini penyakit. Salah satu algoritma yang cukup populer dalam memprediksi kejadian biner seperti keberadaan penyakit jantung adalah algoritma *Logistic Regression*. Algoritma ini mampu memodelkan hubungan antara beberapa variabel independen seperti usia, tekanan darah, kolesterol, dan kebiasaan merokok dengan kemungkinan terjadinya penyakit jantung.

Beberapa penelitian telah menunjukkan efektivitas algoritma *Logistic Regression* dalam memprediksi resiko penyakit jantung. (Kurniawan Saputro et al., 2023) penerapan algoritma Logistic Regression terbukti efektif dalam memprediksi risiko penyakit jantung dengan tingkat akurasi yang cukup tinggi. Namun, terdapat tantangan dalam penanganan data yang hilang, seleksi fitur yang tepat, serta integrasi dengan teknologi medis terkini yang perlu diperhatikan untuk meningkatkan akurasi dan efektivitas deteksi penyakit jantung. Selain itu, penelitian oleh (Pangaribuan et al., 2021) menunjukkan

bahwa algoritma Logistic Regression memiliki tingkat sensitivitas sebesar 88,54% pada data training dan spesifisitas sebesar 87,50% pada data testing. Hal ini menunjukkan bahwa algoritma tersebut cukup andal dalam mendeteksi keberadaan penyakit jantung.

(Sitanggang et al., 2022) juga membandingkan algoritma *Logistic Regression* dengan algoritma K-Nearest Neighbor (K-NN), dan hasilnya menunjukkan bahwa algoritma Logistic Regression menghasilkan akurasi lebih tinggi, yaitu sebesar 88% dibandingkan dengan algoritma K-NN yang hanya mencapai 83%. Hal serupa ditemukan oleh (Widiawati et al., 2024) yang mengembangkan sistem prediksi berbasis website dengan menggunakan algoritma Logistic Regression dan memperoleh akurasi sebesar 90%, dengan presisi sebesar 92%, recall 86%, dan f1-score 89%. Penelitian lain oleh (Azhar & Sari, 2022) juga membuktikan keunggulan algoritma Logistic Regression dengan akurasi sebesar 88,52% mengungguli algoritma lain seperti K-NN dan Random Forest.

Selanjutnya, penelitian oleh (Asyafiyah & Akbar, 2024) menunjukkan bahwa model prediksi penyakit jantung berbasis Logistic Regression menghasilkan akurasi sebesar 86% pada data training dan 88% pada data testing, dengan nilai Area Under Curve (AUC) sebesar 0,95. Nilai AUC ini menunjukkan performa model yang sangat baik dalam mengklasifikasikan data. Adapun penelitian dari (Napiyah & Heristian, 2024) menunjukkan bahwa algoritma Logistic Regression mencapai akurasi sebesar 85% dan direkomendasikan sebagai salah satu metode efektif dalam prediksi penyakit jantung.

Berdasarkan berbagai penelitian tersebut, dapat disimpulkan bahwa algoritma Logistic Regression memiliki potensi besar dalam mendeteksi dan memprediksi risiko penyakit jantung secara akurat. Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma Logistic Regression dalam membangun model prediksi penyakit jantung guna meningkatkan akurasi identifikasi pasien berisiko serta mendukung upaya pencegahan penyakit jantung secara lebih efektif. Dengan model prediksi yang tepat, diharapkan tenaga medis dapat terbantu dalam pengambilan keputusan dan mampu memberikan intervensi yang lebih cepat serta meningkatkan kualitas hidup pasien.

1.2 Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana penerapan algoritma Logistic Regression untuk membangun model prediksi penyakit jantung?
2. Seberapa baik akurasi dan kinerja model dalam memprediksi risiko penyakit jantung?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian yang akan dilakukan adalah sebagai berikut:

1. Menerapkan algoritma Logistic Regression dalam memprediksi kemungkinan seseorang menderita penyakit jantung berdasarkan faktor-faktor risiko yang relevan.
2. Menganalisis efektivitas dan akurasi algoritma Logistic Regression dalam memprediksi penyakit jantung.
3. Mengidentifikasi variabel-variabel independen yang memiliki pengaruh signifikan terhadap prediksi penyakit jantung.
4. Menganalisis performa model Logistic Regression dalam memprediksi penyakit jantung menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score.
5. Menghasilkan model prediksi yang dapat digunakan sebagai alat bantu dalam deteksi dini penyakit jantung guna meningkatkan upaya pencegahan dan pengobatan dini.

1.4 Ruang Lingkup dan Batasan Masalah

1.4.1 Ruang Lingkup

Ruang lingkup penelitian ini mencakup hal-hal berikut:

1. Fokus pada Prediksi Penyakit Jantung
Penelitian ini difokuskan pada penerapan algoritma Logistic Regression untuk memprediksi kemungkinan seseorang mengidap penyakit jantung berdasarkan dataset yang tersedia.
2. Penggunaan Dataset Terstandarisasi
Penelitian ini menggunakan dataset terbuka (seperti Cleveland Heart Disease dataset dari UCI Machine Learning Repository) yang diunduh dari website Kaggle dan telah banyak digunakan dalam penelitian serupa.
3. Fitur yang Digunakan
Fitur-fitur yang digunakan dalam model prediksi meliputi atribut-atribut yang umum dalam diagnosis penyakit jantung, seperti usia, jenis kelamin, tekanan darah, kadar kolesterol, denyut jantung maksimal, dan lain-lain sesuai dengan dataset.
4. Algoritma yang Digunakan
Algoritma utama yang digunakan adalah Logistic Regression, dengan evaluasi menggunakan metrik seperti akurasi, presisi, recall, dan F1-score.
5. Implementasi dengan Python
Proses implementasi algoritma dilakukan menggunakan bahasa pemrograman Python dengan bantuan pustaka seperti *scikit-learn*, *pandas*, dan *matplotlib*.

1.4.2 Batasan Masalah

Agar penelitian lebih terfokus dan dapat diselesaikan dengan baik, maka dibatasi oleh hal-hal berikut:

1. Tidak Mencakup Diagnosa Langsung
Penelitian ini tidak digunakan untuk diagnosis medis secara langsung,

- melainkan hanya sebagai sistem bantu berbasis data untuk memprediksi kemungkinan penyakit jantung.
2. Hanya Menggunakan Algoritma Logistic Regression
Meskipun banyak algoritma lain seperti Decision Tree, Random Forest, atau Neural Network yang dapat digunakan, penelitian ini hanya membahas dan menerapkan Logistic Regression.
 3. Data yang Digunakan Bersifat Sekunder
Data yang digunakan dalam penelitian ini merupakan data sekunder dari sumber terbuka, bukan data primer dari rumah sakit atau klinik lokal.
 4. Validasi Model Secara Statistik
Validasi model hanya dilakukan secara statistik menggunakan metode train-test split, bukan melalui uji klinis atau konsultasi dengan tenaga medis profesional.
 5. Tidak Mencakup Pengembangan Aplikasi Berbasis Web atau Mobile
Penelitian ini hanya berfokus pada pembangunan dan evaluasi model prediksi, tidak mencakup pembuatan sistem aplikasi akhir yang dapat digunakan langsung oleh pengguna umum.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian yang akan dilakukan adalah sebagai berikut:

1. Memberikan contoh nyata penerapan algoritma Logistic Regression di bidang kesehatan sehingga dapat dimanfaatkan sebagai acuan dalam pengembangan sistem pendukung keputusan untuk deteksi penyakit jantung.
2. Memberikan informasi mengenai keandalan algoritma Logistic Regression sebagai metode prediksi, sehingga dapat dijadikan dasar pemilihan algoritma yang tepat pada penelitian maupun implementasi nyata.
3. Membantu pihak medis maupun masyarakat memahami faktor risiko yang paling berpengaruh terhadap penyakit jantung sehingga dapat dilakukan tindakan pencegahan yang lebih tepat sasaran.
4. Memberikan gambaran yang lebih komprehensif mengenai kelebihan dan keterbatasan model Logistic Regression dalam prediksi penyakit jantung, sehingga hasil penelitian dapat menjadi dasar pengembangan model yang lebih optimal di masa mendatang.
5. Menyediakan sistem prediksi yang praktis dan bermanfaat sebagai pendukung keputusan bagi tenaga medis maupun masyarakat untuk mendeteksi risiko penyakit jantung lebih awal, sehingga dapat menekan angka kematian akibat keterlambatan penanganan.