

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi informasi berkembang sangat penting dalam proses kehidupan manusia. Seiring dengan perkembangan ini, jumlah data yang dihasilkan pun semakin besar dan terus bertambah. Sebelum data dapat digunakan secara efektif dan memiliki nilai yang valid diperlukan metode untuk memproses, menyimpan, serta mengelolanya dengan baik. Salah satu pendekatan yang dapat digunakan untuk menangani hal tersebut adalah pemanfaatan teknologi basis data yang dikenal dengan istilah data mining.

Data mining adalah proses pengolahan data yang memanfaatkan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstrak serta mengidentifikasi informasi yang bermanfaat dan pengetahuan yang relevan dari kumpulan data berskala besar. Dalam data mining, terdapat informasi baru, salah satunya adalah klasifikasi.

Klasifikasi merupakan proses pengelompokan atau pemisahan data berdasarkan karakteristik tertentu. Fungsi utama dari klasifikasi adalah menjelaskan dan membedakan suatu konsep atau kategori data dengan tujuan menentukan perkiraan kelas dari objek yang belum memiliki label atau membagi suatu data ke dalam kelompok-kelompok yang sesuai (Dinata et al., 2020). Klasifikasi dapat digunakan di berbagai bidang yaitu bidang perbankan, perdagangan, industri, dan juga kesehatan. Di bidang kesehatan, metode klasifikasi banyak digunakan dalam proses diagnosis penyakit, evaluasi faktor resiko, dan pengambilan keputusan teknis.

Kesehatan adalah hal yang paling berharga bagi setiap individu, karena siapa pun dapat mengalami gangguan kesehatan. Oleh karena itu, penting bagi kita untuk menjaga kesehatan dengan baik agar terhindar dari berbagai penyakit, termasuk penyakit jantung. Jantung merupakan bagian organ vital yang paling kuat, karena fungsi utamanya adalah memompa darah melalui aliran darah ke seluruh bagian

tubuh (Muharni & Andriyanto, 2021). Penyakit jantung atau kardiovaskular mencakup berbagai kondisi yang terjadinya penyempitan atau penyumbatan pembuluh darah yang menjadi sebab stroke, nyeri dada atau angina, dan serangan jantung. Gejala yang umum dialami oleh penderita penyakit jantung meliputi keluhan nyeri dada yang terasa seperti tertimpa beban, terjepit, tertekan, terbakar, atau bahkan seperti tercekik. Rasa nyeri ini biasanya muncul di bagian tengah dada dan dapat menjalar ke lengan kiri, leher, hingga ke punggung. Selain itu, gejala lainnya meliputi sesak nafas, jantung berdebar-debar, kecemasan, pusing berkepanjangan, keringat dingin, tubuh terasa lemas, hingga pingsan. Faktor risiko utama penyakit ini adalah pola hidup yang tidak sehat, termasuk pola makan yang buruk dan kurangnya aktivitas fisik.

Berdasarkan data dari *World Health Organization* (WHO) tahun 2022, penyakit jantung merupakan penyakit mematikan tertinggi nomor satu di dunia, terutama menyerang orang dewasa dan lanjut usia. Hingga saat ini, tercatat sekitar 17,9 juta kematian setiap tahun akibat penyakit ini, dengan gagal jantung menjadi faktor utama yang berkontribusi terhadap 85% kematian pada penderita penyakit kardiovaskular. Sebagian besar kasus kematian akibat penyakit jantung terjadi di negara-negara berpenghasilan rendah hingga menengah, termasuk Indonesia. Di Asia tercatat angka kematian akibat penyakit jantung, yaitu sekitar 712,1 ribu jiwa. Sementara itu, Indonesia menempati posisi kedua di kawasan Asia Tenggara dengan jumlah kasus kematian akibat penyakit jantung mencapai 371 ribu jiwa (Harisa et al., 2020).

Melihat tingginya angka kasus penyakit jantung dan dampaknya yang signifikan terhadap kesehatan masyarakat, diperlukan metode deteksi dini yang lebih akurat agar langkah pencegahan dan pengobatan dapat dilakukan dengan lebih cepat dan efektif. Salah satu pendekatan yang dapat meningkatkan keakuratan diagnosis adalah dengan menerapkan data mining dalam klasifikasi penyakit jantung berdasarkan data rekam medis. Data rekam medis mengandung informasi penting tentang kondisi kesehatan pasien, yang apabila dianalisis menggunakan metode yang tepat, dapat membantu tenaga medis dalam mengenali pola serta faktor risiko penyakit jantung dengan lebih optimal.

Salah satu metode yang telah terbukti efektif dalam klasifikasi data medis adalah *Logistic Regression*. *Logistic Regression* merupakan salah satu metode analisis statistik yang digunakan untuk memprediksi kemungkinan terjadinya suatu kejadian berdasarkan sekumpulan variabel independen. Metode ini banyak digunakan dalam klasifikasi biner, di mana hasil prediksi terbagi menjadi dua kategori, seperti "positif" atau "negatif", "beresiko" atau "tidak beresiko", serta "ya" atau "tidak", dan sebagainya.

Berdasarkan penelitian sebelumnya menggunakan metode logistic regression yang dikutip dalam jurnal Ermawati, Risnawati Ibna, dan Baso Agung Kurniawan pada tahun 2023 dengan judul "Klasifikasi Penderita Anemia Menggunakan Metode Regresi Logistik". Metode regresi logistik untuk mengklasifikasikan anemia pada remaja putri, dengan variabel signifikan berupa ferritin serum dan transferrin receptor (STfr). Model yang dihasilkan memiliki akurasi 79,23%, presisi 80,18%, dan recall 96,39%, menunjukkan kinerja baik dalam mengidentifikasi anemia (Ibna & Agung Kurniawan, 2023).

Adapun penelitian lain oleh Gita Rohma Utami Asyafiiyah, Ronny Makhfuddin Akbar yang berjudul "Prediksi Pasien Terindikasi Penyakit Jantung Menggunakan Metode Logistic Regression". Metode Logistic Regression mampu memprediksi risiko penyakit jantung dengan akurasi tinggi, yaitu 86% pada data pelatihan dan 88% pada data pengujian, tanpa mengalami overfitting atau underfitting yang signifikan. Evaluasi menggunakan confusion matrix menunjukkan model memiliki True Positive (TP) sebesar 90,48%, True Negative (TN) sebesar 85,39%, serta False Negative (FN) yang rendah, yaitu 9,52%. Sementara itu, ROC curve menghasilkan nilai AUC 0,95, menandakan performa model yang sangat baik dalam membedakan pasien dengan atau tanpa penyakit jantung. Penelitian ini membuktikan bahwa metode Logistic Regression efektif dalam memprediksi risiko penyakit jantung dan dapat dimanfaatkan sebagai alat bantu tenaga medis untuk deteksi dini dan pencegahan yang lebih optimal (Rohma et al., 2024)

Kemudian penelitian lain oleh Zaenal Mutaqin, Chaerur Rozikin, Yusrizal Anastya Tomo dengan judul "Klasifikasi Penyakit Diabetes Menggunakan

Algoritma Logistic Regression”. Metode Logistic Regression terbukti efektif dalam mengklasifikasikan pasien yang berisiko mengalami diabetes dengan tingkat akurasi yang cukup tinggi. Penelitian ini menguji tiga model klasifikasi dengan variasi dalam pembagian data pelatihan dan pengujian. Berdasarkan hasil evaluasi, model pertama menunjukkan akurasi tertinggi sebesar 81%, dengan 96 hasil True Positive (TP), 9 False Positive (FP), 21 False Negative (FN), dan 28 True Negative (TN). Sementara itu, model kedua memiliki akurasi 80%, dan model ketiga mencapai 75% (Mutaqin et al., 2024).

Berdasarkan uraian yang telah disampaikan dari berbagai referensi yang diperoleh, penulis tertarik untuk melakukan penelitian dengan menggunakan metode logistic regression, mengingat kemampuan metode ini dalam akurasi tinggi dalam klasifikasi data. Oleh karena itu, penulis memilih judul penelitian “Klasifikasi Penyakit Jantung Berdasarkan Data Rekam Medis Menggunakan Metode Logistic Regression”.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang diuraikan diatas, maka permasalahan yang penulis rumuskan adalah:

1. Bagaimana menerapkan metode logistic regression dalam mengklasifikasikan penyakit jantung?
2. Bagaimana besar tingkat hasil akurasi metode logistic regression pada proses pengklasifikasian penyakit jantung?

1.3 Tujuan Penelitian

Dari rumusan masalah yang telah dirumuskan, tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui metode Logistic Regression dalam melakukan klasifikasi penyakit jantung
2. Untuk memastikan bahwa penerapan logistic regression dalam klasifikasi penyakit jantung digunakan dengan mudah dan bermanfaat oleh pengguna.

3. Untuk mengetahui keefektifitasan logistic regression pada klasifikasi penyakit jantung.

1.4 Manfaat Penelitian

Penulis berharap penelitian ini dapat memberikan manfaat di kemudian hari. Adapun manfaat yang diharapkan penulis adalah sebagai berikut:

1. Bagi Peneliti:
 - a. Menerapkan ilmu yang telah diperoleh selama perkuliahan dalam penelitian berbasis data mining.
 - b. Menambah wawasan dan pemahaman lebih dalam mengenai penerapan Logistic Regression dalam klasifikasi penyakit jantung.
 - c. Memenuhi salah satu syarat dalam menyelesaikan program studi S1 Teknik Informatika Universitas Malikussaleh.
2. Bagi Pembaca:
 - a. Membantu pembaca memahami cara kerja algoritma Logistic Regression dalam klasifikasi penyakit jantung berdasarkan data rekam medis.
 - b. Menjadi referensi bagi mahasiswa atau peneliti yang ingin mengembangkan penelitian serupa, khususnya dalam bidang klasifikasi penyakit menggunakan metode machine learning.

1.5 Ruang Lingkup dan Batasan Penelitian

Agar tujuan penelitian ini tercapai, maka penelitian perlu dibatasi. Adapun batasan penelitian yang dibuat penulis adalah:

1. Sumber data penelitian ini didapatkan dari penelitian (studi) terdahulu.
2. Metode yang digunakan dalam penelitian ini adalah logistic regression.
3. Data yang digunakan sebagai objek penelitian adalah tahun 2021-2023.
4. Variabel yang yang digunakan yaitu usia, jenis kelamin, HR (*Heart Rate*), RR (*Respiratory Rate*) dan nyeri dada.
5. Hasil klasifikasi dibagi menjadi dua yaitu pasien beresiko penyakit jantung koroner dan pasien tidak beresiko penyakit jantung koroner.