

ABSTRAK

Klasifikasi merupakan salah satu teknik dalam data mining yang digunakan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan karakteristik masing-masing objek data. Penelitian ini menggunakan algoritma Logistic Regression untuk mengklasifikasikan risiko penyakit jantung berdasarkan data rekam medis pasien. Penyakit jantung merupakan kondisi medis serius dan menjadi salah satu penyebab kematian tertinggi secara global, termasuk di Indonesia, serta memberikan beban besar terhadap biaya layanan kesehatan. Oleh karena itu, deteksi dini sangat penting untuk mendukung pencegahan dan pengobatan yang lebih efektif. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi penyakit jantung guna membantu mengidentifikasi pasien yang berisiko secara cepat dan akurat. Dataset yang digunakan terdiri dari 100 data pasien tahun 2021 hingga 2023, dengan 80 data untuk pelatihan dan 20 untuk pengujian. Atribut yang digunakan meliputi jenis kelamin, usia, tekanan darah, heart rate (HR), respiratory rate (RR), dan nyeri dada. Hasil evaluasi menunjukkan bahwa algoritma Logistic Regression memberikan performa klasifikasi yang baik, dengan akurasi 95%, precision 95,83%, recall 94,44%, dan F1-score 94,89%. Hasil ini menunjukkan bahwa model dapat digunakan sebagai alat bantu dalam mendeteksi risiko penyakit jantung secara dini dan efektif.

Kata kunci: Data Mining, Klasifikasi, Logistic Regression, Penyakit Jantung.

ABSTRACT

Classification is a data mining technique used to group data into specific categories based on the characteristics of each data object. This study applies the Logistic Regression algorithm to classify the risk of heart disease based on patient medical records. Heart disease is a serious medical condition and one of the leading causes of death globally, including in Indonesia, and contributes significantly to healthcare costs. Early detection is therefore essential to support effective treatment and prevention. The purpose of this research is to develop a heart disease classification system to help identify patients at risk quickly and accurately. The dataset used consists of 100 patient records from 2021 to 2023, with 80 used for training and 20 for testing. The attributes include gender, age, blood pressure, heart rate (HR), respiratory rate (RR), and chest pain. Evaluation of the model shows that Logistic Regression performed well in classifying risk, with an accuracy of 95%, precision of 95.83%, recall of 94.44%, and an F1-score of 94.89%. These results indicate that the model is reliable for early detection of heart disease risk using structured medical data and can support decision-making in clinical settings.

Keywords: *Data Mining, Classification, Logistic Regression, Heart Disease*