

ABSTRAK

Penyakit kardiovaskular masih menjadi salah satu penyebab kematian tertinggi di dunia, dan serangan jantung merupakan salah satu komplikasi yang banyak dipengaruhi oleh hipertensi dan diabetes melitus. Pada pasien dengan kedua kondisi tersebut, deteksi dini risiko serangan jantung menjadi penting untuk mendukung upaya pencegahan. Dalam konteks ini, machine learning dapat dimanfaatkan untuk mempelajari pola data secara lebih fleksibel dibandingkan pendekatan konvensional. Oleh karena itu, penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja beberapa model machine learning dalam prediksi risiko serangan jantung pada pasien hipertensi dan diabetes melitus.

Penelitian ini menggunakan data sekunder berformat CSV dari platform Kaggle yang terdiri atas 158.355 baris data dan 22 kolom, dengan 21 variabel input dan 1 variabel target, yaitu heart_attack. Metode penelitian meliputi pembacaan dataset, pengecekan missing values, encoding data kategorikal, normalisasi menggunakan MinMaxScaler, pembagian data training dan testing dengan rasio 80:20, penanganan ketidakseimbangan kelas menggunakan SMOTE, pelatihan model, tuning hyperparameter, pengujian model, analisis komparasi, penentuan model terbaik, serta implementasi prediksi. Model yang dibandingkan adalah Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), dan Logistic Regression. Evaluasi dilakukan menggunakan metrik accuracy, precision, recall, F1-score, dan AUC-ROC.

Hasil penelitian menunjukkan bahwa dataset tidak memiliki missing values, seluruh atribut kategorikal berhasil diubah ke bentuk numerik, dan distribusi kelas pada data training berhasil diseimbangkan dari 59,9% : 40,1% menjadi 50,0% : 50,0% setelah penerapan SMOTE. Berdasarkan hasil evaluasi, Random Forest memperoleh accuracy 0,7193, precision 0,6412, recall 0,6815, F1-score 0,6607, dan AUC 0,8015, sehingga ditetapkan sebagai model terbaik. Sementara itu, Logistic Regression menghasilkan F1-score tertinggi sebesar 0,6618 dan KNN menghasilkan recall tertinggi sebesar 0,7040. Dengan demikian, penelitian ini menunjukkan bahwa Random Forest merupakan model yang paling optimal untuk prediksi risiko serangan jantung pada pasien hipertensi dan diabetes melitus, meskipun hasil prediksi yang diperoleh tidak dimaksudkan sebagai alat diagnosis medis.

Kata kunci: machine learning, prediksi risiko, serangan jantung, hipertensi, diabetes melitus, Random Forest.