

ABSTRACT

Heart disease is one of the cardiovascular diseases (CVD) that is a leading global cause of death, contributing to approximately 32% of all global deaths. This can be mitigated by predicting heart disease. Clinical Decision Support Systems (CDSS) can be used to assist doctors in assessing patients' conditions. There has been extensive research on CDSS for predicting heart disease; however, very few have incorporated outlier detection and data balancing before making predictions.

Therefore, this research aims to enhance the accuracy of heart disease prediction by integrating Local Outlier Factor (LOF) for outlier detection, SMOTE-ENN for data balancing, and the XGBoost algorithm for heart disease prediction. It is expected that the improved accuracy of the heart disease prediction model can be applied to reduce heart disease-related mortality.

The Statlog dataset is used to create the model and compare it with models from previous research, such as LR, hybrid LR and linear models, and HDPM DBSCAN, SMOTE-ENN, and XGBoost. The results of the proposed heart disease prediction model which is 97.71% accuracy demonstrate higher than previous studies.

Keywords : Heart disease, machine learning, Local Outlier Factor (LOF), SMOTE-ENN, XGBoost.

INTISARI

Penyakit jantung adalah salah satu penyakit kardiovaskular (CVD) yang menjadi penyebab utama kematian secara global dan berkontribusi sekitar 32% dari semua kematian global. Hal ini dapat dimitigasi dengan memprediksi penyakit jantung. Sistem pendukung keputusan klinis (SPKK) dapat digunakan untuk membantu dokter dalam menilai kondisi pasien. Dewasa ini telah banyak penelitian SPKK untuk memprediksi penyakit jantung, akan tetapi masih sangat sedikit yang menggunakan deteksi data *outlier* dan menyeimbangkan data sebelum dilakukan prediksi.

Oleh karena itu penelitian ini bertujuan untuk meningkatkan akurasi prediksi penyakit jantung dengan mengintegrasikan *Local Outlier Factor* (LOF) untuk deteksi data *outlier*, SMOTE-ENN untuk menyeimbangkan data, dan algoritma XGBoost untuk memprediksi penyakit jantung. Diharapkan dengan meningkatnya akurasi model prediksi penyakit jantung dapat diterapkan sehingga mengurangi angka kematian akibat penyakit jantung.

Dataset Statlog digunakan untuk membuat model dan membandingkan dengan model dari penelitian sebelumnya. Model penelitian sebelumnya seperti LR; hibrida LR dan *linear model*; dan HDPM DBSCAN, SMOTE-ENN, dan XGBoost. Hasil dari model prediksi penyakit jantung yang diusulkan memiliki akurasi yaitu 97,71% yang mana lebih tinggi dari penelitian-penelitian sebelumnya.

Kata Kunci : Penyakit jantung, *machine learning*, *Local Outlier Factor* (LOF), SMOTE-ENN, XGBoost.