

ABSTRAK

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia yang membutuhkan sistem deteksi dini yang akurat dan efisien. Penelitian kuantitatif dengan pendekatan eksperimental komputasional ini bertujuan untuk mengoptimalkan performa model klasifikasi penyakit jantung biner (sehat atau sakit). Optimasi dilakukan melalui integrasi teknik reduksi dimensi CFS pada algoritma *ensemble Random Forest* menggunakan UCI *Heart Disease Dataset*. CFS bekerja dengan mempertahankan fitur yang berkorelasi tinggi terhadap kelas target sekaligus mengeliminasi fitur yang redundan. Sebelum pemodelan, data melalui tahap pra-pemrosesan yang meliputi penanganan *missing values* menggunakan imputasi modus serta standardisasi metrik numerik dengan *Z-score Standardization*. Hasil eksperimen menunjukkan bahwa penerapan CFS berhasil mereduksi dimensi data sebesar 46%, yaitu menyaring 13 atribut awal menjadi 7 atribut paling dominan (*age*, *cp*, *thalach*, *oldpeak*, *slope*, *ca*, dan *thal*). Pengurangan dimensi ini secara signifikan menurunkan kompleksitas model dan memitigasi risiko *overfitting*. Model optimasi (CFS + *Random Forest*) berhasil mencapai tingkat akurasi sebesar 88,1%, Presisi 87%, *Recall* 89%, dan *F1-Score* 88%. Performa ini menunjukkan peningkatan yang signifikan dibandingkan dengan model *baseline* tanpa seleksi fitur yang hanya memiliki akurasi 83,5%, Presisi 82%, *Recall* 85%, dan *F1-Score* 83,4%. Nilai *Recall* yang tinggi (89%) dengan hanya 9 sampel *False Negative* menegaskan bahwa model ini sangat sensitif dan *robust* untuk digunakan sebagai landasan saintifik dalam pengembangan Sistem Pendukung Keputusan Klinis (*Clinical Decision Support System*) demi menekan angka mortalitas akibat penyakit kardiovaskular.

Kata Kunci: Penyakit Jantung, *Correlation-based Feature Selection*, *Random Forest*, Klasifikasi, *Clinical Decision Support System*.