

ABSTRAK

Penyakit kardiovaskular (CVD) merupakan penyebab utama kematian di dunia dan sering kali berkembang tanpa gejala awal yang jelas. Deteksi dini terhadap risiko penyakit ini menjadi krusial untuk menurunkan angka kematian dan meningkatkan efektivitas intervensi klinis. Penelitian ini bertujuan untuk membangun model prediksi risiko penyakit jantung berdasarkan data *Framingham Heart Study* menggunakan kombinasi metode *Adaptive Synthetic Sampling* (ADASYN) dan algoritma *Extreme Gradient Boosting* (XGBoost). ADASYN digunakan untuk mengatasi ketidakseimbangan kelas pada data, sedangkan XGBoost dipilih karena kemampuannya dalam menangani data tabular dan menghasilkan model prediktif yang akurat dan efisien. Penelitian ini melibatkan enam skenario eksperimen dengan variasi pada penggunaan ADASYN, tuning hyperparameter, dan seleksi fitur. Evaluasi model dilakukan menggunakan metrik *accuracy*, *recall*, *precision*, *F1-score*, dan *area under the curve* (AUC), dengan fokus pada performa kelas minoritas (pasien berisiko). Hasil terbaik diperoleh pada skenario yang menggunakan ADASYN, *tuning*, dan seleksi fitur, dengan *recall* mencapai 25%, *F1-score* sebesar 26%, dan AUC sebesar 64.3%. Peningkatan ini jauh lebih baik dibandingkan model *baseline* yang hanya memiliki *recall* 7% dan *F1-score* 10%. Meskipun nilai metrik masih tergolong rendah secara absolut, peningkatan sensitivitas terhadap pasien berisiko memiliki makna besar dalam konteks medis. Oleh karena itu, kombinasi ADASYN dan XGBoost yang digunakan dalam penelitian ini dinilai mampu memberikan kontribusi dalam mendukung pengembangan sistem deteksi dini risiko penyakit jantung secara lebih adil, sensitif, dan aplikatif di lingkungan klinis.

Kata Kunci: prediksi penyakit jantung, ADASYN, XGBoost, data tidak seimbang, seleksi fitur, *framingham heart study*, *machine learning*.