

ABSTRAK

Penyakit jantung tetap menjadi salah satu faktor utama penyebab kematian di berbagai belahan dunia, sehingga diperlukan diagnosis dini yang akurat untuk mengurangi risiko yang ditimbulkan. Kemajuan teknologi machine learning memberikan peluang baru untuk membantu tenaga medis dalam memprediksi penyakit jantung secara lebih efisien dan tepat. Kajian ini mengarah pada evaluasi dan perbandingan terhadap kinerja dua algoritma pembelajaran terawasi yang populer, yaitu *Random Forest* dan *Gradient Boosting*, dalam klasifikasi penyakit jantung. Dataset berisikan 1.000 baris data dengan sejumlah fitur yang merepresentasikan berbagai faktor risiko penyakit jantung. Penilaian kinerja dilakukan dengan memanfaatkan metrik seperti akurasi, presisi, *recall* dan *f1-score*. Dari hasil analisis diperoleh bahwa *Random Forest* unggul dibandingkan *Gradient Boosting* dalam seluruh metrik evaluasi. *Random Forest* memperoleh akurasi sebesar 99,5%, sementara *Gradient Boosting* memperoleh 98,5%. Selain itu, *Random Forest* mencapai nilai sempurna (100%) pada presisi kelas 0, *recall* kelas 1, dan *F1-score* kelas 1, menunjukkan kemampuannya yang tinggi dalam klasifikasi penyakit jantung. Model yang dikembangkan ini memiliki potensi besar untuk diterapkan sebagai alat bantu pengambilan keputusan dalam sistem layanan kesehatan, terutama pada tahap skrining awal dan penilaian risiko pasien. Dengan mengidentifikasi pola dan fitur kunci yang berhubungan dengan penyakit jantung, model ini dapat membantu tenaga kesehatan dalam memberikan hasil klinis yang lebih cepat juga tepat sasaran.

Kata kunci : penyakit jantung, klasifikasi, *random forest*, *gradient boosting*.

ABSTRACT

Cardiovascular disease still accounts for a major portion of deaths worldwide, emphasizing the need for accurate and early diagnosis to reduce associated risks. Advances in machine learning provide opportunities to assist medical professionals in predicting heart disease more efficiently. This study evaluates and compares the performance of two popular supervised learning algorithms, Random Forest and Gradient Boosting, for heart disease classification. Using a dataset comprised of 1000 records highlighting various heart disease risk indicators, the models were developed and assessed. Their performance was determined based on accuracy, precision, recall, and f1-score metrics. The findings indicate that Random Forest consistently outperforms Gradient Boosting across all evaluation metrics. Specifically, Random Forest achieved an accuracy of 99.5%, surpassing Gradient Boosting's 98.5%. Moreover, Random Forest achieved perfect scores (100%) in class 0 precision, class 1 recall, and class 1 F1-score, demonstrating its robustness in handling heart disease classification. The developed model presents strong potential as a decision-support tool in healthcare settings, especially in early screening and patient risk assessment. By identifying key features and patterns associated with heart disease, such models can support healthcare professionals in making quicker and more targeted clinical decisions, ultimately contributing to improved patient outcomes.

Keywords: heart disease, classification, random forest, gradient boosting.



KARAWANG