

ABSTRAK

Penelitian ini bertujuan untuk mengevaluasi performa algoritma *Logistic Regression*, *Decision Tree*, dan *LightGBM* dalam memprediksi konflik klasifikasi pada varian genetik. Dengan pertumbuhan teknologi *genomics* yang pesat, pengumpulan data genetik dalam skala besar telah menjadi lebih mudah, memungkinkan analisis mendalam untuk menilai risiko penyakit dan respons terhadap terapi medis. Masalah yang dihadapi adalah ketidakpastian dalam mengidentifikasi varian genetik yang dapat memiliki klasifikasi yang bertentangan, yang dapat mempengaruhi keputusan klinis dan penelitian genetik. Solusi yang diusulkan adalah menggunakan tiga model *machine learning* yang berbeda untuk memprediksi kemungkinan konflik klasifikasi, dengan evaluasi menggunakan metrik akurasi, *precision*, *recall*, *F1-score*, dan *specificity* melalui teknik *cross-validation*. Hasil dari penelitian ini diharapkan memberikan wawasan penting tentang efektivitas pendekatan *machine learning* dalam analisis data genetik, serta memberikan rekomendasi untuk aplikasi klinis dan arah penelitian mendatang dalam *genomics*.

Kata Kunci : *Logistic Regression*, *Decision Tree*, *LightGBM*, klasifikasi varian genetik, prediksi konflik, *machine learning*, *genomics*, evaluasi model

ABSTRACT

This study aims to evaluate the performance of Logistic Regression, Decision Tree, and LightGBM algorithms in predicting classification conflicts in genetic variants. With the rapid growth of genomic technology, large-scale genetic data collection has become more accessible, enabling in-depth analysis for assessing disease risks and responses to medical therapies. The problem addressed is the uncertainty in identifying genetic variants that may have conflicting classifications, which can influence clinical decisions and genetic research. The proposed solution involves using three different machine learning models to predict potential classification conflicts, evaluated using accuracy, precision, recall, F1-score, and specificity metrics through cross-validation techniques. The findings of this research are expected to provide critical insights into the effectiveness of machine learning approaches in genetic data analysis and offer recommendations for clinical applications and future research directions in genomics.

Keywords : Logistic Regression, Decision Tree, LightGBM, genetic variant classification, conflict prediction, machine learning, genomics, model evaluation