

COMPARATIVE ANALYSIS OF RANDOM FOREST AND GRADIENT BOOSTING FOR EARLY DETECTION OF DENGUE FEVER IN REGIONAL HOSPITAL

By

Made Dwi Aprillia Kusuma Wiryani, NIM. 22115091106

Department of Informatics Engineering

Information Systems Program

ABSTRACT

Dengue Fever (DF) is a disease caused by the dengue virus (DENV), primarily transmitted through the bites of infected mosquitoes. Dengue surged globally in 2024 with 7.6 million dengue cases reported by the World Health Organization (WHO). Dengue fever shares overlapping symptoms with other febrile conditions, such as febris, influenza, chikungunya, etc., commonly presenting with high body temperature, making early detection difficult without prior laboratory examination. Machine learning offers a way to support early detection through classification. This study compares two algorithms, Random Forest and Gradient Boosting, both well-known for classification tasks, evaluated using accuracy, precision, recall, and F1-score. Recall is prioritized as the primary metric to minimize false negatives in disease classification. The dataset was obtained from RSUD Buleleng through an ethical procedure, consisting of 212 entries, with 211 cleaned records and 11 selected clinical features used in modelling. Stratified 5-fold cross-validation and GridSearchCV applied to ensure both classes preserved in training and testing. Under stratified 5-fold cross-validation, RF maintained its advantage with a mean CV recall of 87.32%, slightly higher than GB's mean CV recall of 81.76%, leading to RF being selected for final model deployment, implemented via a React.js and Flask API.

Keywords: Classification, Dengue, Gradient Boosting, Random Forest, Recall

ANALISIS PERBANDINGAN RANDOM FOREST DAN GRADIENT BOOSTING UNTUK DETEKSI DINI DEMAM BERDARAH DENGUE PADA RUMAH SAKIT UMUM DAERAH

Oleh

Made Dwi Aprillia Kusuma Wiryani, NIM. 22115091106

Jurusan Teknik Informatika

Program Studi Sistem Informasi

ABSTRAK

Demam Berdarah Dengue (DBD) adalah penyakit yang disebabkan oleh virus dengue (DENV), yang terutama ditularkan melalui gigitan nyamuk yang terinfeksi. Kasus dengue melonjak secara global pada tahun 2024 dengan 7,6 juta kasus dilaporkan oleh Organisasi Kesehatan Dunia (WHO). Demam berdarah memiliki gejala yang tumpang tindih dengan kondisi demam lainnya, seperti demam, influenza, chikungunya, dll., yang umumnya ditandai dengan suhu tubuh tinggi, sehingga menyulitkan deteksi dini tanpa pemeriksaan laboratorium sebelumnya. Machine learning menawarkan cara untuk mendukung deteksi dini melalui klasifikasi. Penelitian ini membandingkan dua algoritma, Random Forest dan Gradient Boosting, yang keduanya dikenal dalam tugas klasifikasi, dievaluasi menggunakan akurasi, presisi, recall, dan F1-score. Recall diprioritaskan sebagai metrik utama untuk meminimalkan false negative dalam klasifikasi penyakit. Dataset diperoleh dari RSUD Buleleng melalui prosedur etis, terdiri dari 212 entri, dengan 211 data bersih dan 11 fitur klinis yang dipilih digunakan dalam pemodelan. Stratified 5-fold cross-validation dan GridSearchCV diterapkan untuk memastikan kedua kelas terwakili dalam data pelatihan dan pengujian. Dalam stratified 5-fold cross-validation, RF mempertahankan keunggulannya dengan mean CV recall sebesar 87,32%, sedikit lebih tinggi dibandingkan mean CV recall GB sebesar 81,76%, sehingga RF dipilih untuk penerapan model akhir yang diimplementasikan melalui React.js dan Flask API.

Kata Kunci: Klasifikasi, Dengue, Gradient Boosting, Random Forest, Recall