

DEVELOPING AN APPLICATION FOR EARLY DETECTION OF DENGUE INFECTION USING SMOTE AND THE DECISION TREE ALGORITHM

Arifa Alayya Firajabi

*Program Studi Teknik Komputer, Fakultas Sains dan Teknologi
Universitas Teknologi Yogyakarta
Jl. Ringroad Utara Jombor Sleman Yogyakarta
E-mail : alayyaarifa@gmail.com*

ABSTRACT

Dengue Hemorrhagic Fever (DHF) remains a significant public health concern in Indonesia, where it is endemic and frequently triggers Extraordinary Events (KLB) across various regions. At the Mlati II Community Health Center in Sleman Regency, a primary challenge in the treatment and early diagnosis of dengue is the imbalance in patient medical record data, with the majority class vastly outnumbering the minority class. This imbalance can cause the classification model to become biased and less effective at detecting cases within the minority class. This study aims to develop a web-based early-detection application that integrates the Synthetic Minority Over-sampling Technique (SMOTE) to balance data distribution and employs the Decision Tree algorithm as the primary classification method. Model optimization was conducted by testing 10 experimental scenarios and applying feature selection to create a more efficient, stable, and interpretable model. This study utilized 281 valid patient medical records, with initial features including Platelet Count (AT), Hematocrit (HT), and serological test results comprising NS1, IgG, and IgM. Based on comprehensive testing, Scenario S3, employing a 70:30 training-to-test data ratio and the SMOTE technique without normalization, was selected as the optimal model, achieving an F1 score of 89.36% and an accuracy of 90.59% on the test data. Feature importance analysis revealed that the NS1 antigen parameter was the most influential feature, with a score of 0.6379, and served as the root node in the decision tree. In contrast, the IgG and IgM features were excluded from the model due to their zero contribution and redundancy. The final optimized model, incorporating three primary features, AT, HT, and NS1, was implemented as a web-based application using the Streamlit framework. Overall, the developed system has demonstrated itself to be an objective, efficient, and quantifiable clinical decision-making tool, aiding medical personnel in conducting initial screening of dengue patients in a more systematic, data-driven manner.

Keywords: Dengue, Decision Tree, SMOTE, Streamlit