

ABSTRAK

Anemia defisiensi besi (ADB) merupakan salah satu kondisi kesehatan yang dapat memengaruhi kualitas hidup dan produktivitas masyarakat. Parameter pemeriksaan darah seperti hemoglobin (HGB), Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), dan Mean Corpuscular Hemoglobin Concentration (MCHC) menghasilkan data numerik yang berpotensi dianalisis menggunakan pendekatan *machine learning* untuk membantu proses klasifikasi kondisi anemia. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi anemia defisiensi besi menggunakan algoritma *Random Forest* berdasarkan parameter pemeriksaan darah.

Dataset yang digunakan merupakan dataset sekunder yang diperoleh dari Kaggle dengan jumlah 15.300 sampel dan lima fitur prediktor. Selanjutnya, data dibagi menjadi 80% data latih dan 20% data uji. Proses pemodelan dilakukan menggunakan algoritma *Random Forest* dengan penyesuaian parameter (hyperparameter tuning) secara manual untuk memperoleh konfigurasi model terbaik.

Hasil pengujian menunjukkan bahwa model terbaik diperoleh pada konfigurasi $n_estimators = 50$ dan $max_depth = 5$ dengan skenario pembagian data 80:20. Model tersebut menghasilkan akurasi sebesar 92,48%, precision sebesar 0,81, recall sebesar 0,94, F1-score sebesar 0,87, serta nilai Area Under Curve (AUC) sebesar 0,98, yang menunjukkan kemampuan klasifikasi yang sangat baik dalam membedakan kelas ADB dan Tidak-ADB. Model kemudian diimplementasikan dalam aplikasi berbasis web dan diuji menggunakan *black box* testing serta User Acceptance Test (UAT). Sistem yang dikembangkan berfungsi sebagai alat bantu klasifikasi berbasis data dan tidak dimaksudkan sebagai pengganti diagnosis medis.

Kata kunci: Anemia Defisiensi Besi, *Random Forest*, *Machine Learning*, Sistem Klasifikasi, Evaluasi Model.

ABSTRACT

Iron Deficiency Anemia (IDA) is a health condition that can affect the quality of life and productivity of the population. Blood examination parameters such as hemoglobin (HGB), Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), and Mean Corpuscular Hemoglobin Concentration (MCHC) produce numerical data that can be analyzed using machine learning approaches to assist in the classification of anemia conditions. This study aims to develop a classification system for iron deficiency anemia using the Random Forest algorithm based on blood examination parameters.

The dataset used in this study is a secondary dataset obtained from Kaggle consisting of 15,300 samples with five predictor features. The data underwent exploration and cleaning processes before being used in the modeling stage. Furthermore, the dataset was divided into 80% training data and 20% testing data. The modeling process was performed using the Random Forest algorithm with manual hyperparameter tuning to obtain the best model configuration.

The results show that the best model configuration was achieved with $n_estimators = 50$ and $max_depth = 5$ using an 80:20 data split. The model achieved an accuracy of 92.48%, precision of 0.81, recall of 0.94, F1-score of 0.87, and an Area Under the Curve (AUC) of 0.98, indicating a very good ability to distinguish between IDA and non-IDA classes. The best model was then implemented in a web-based application and evaluated using black box testing and User Acceptance Test (UAT). The developed system functions as a data-based classification support tool and is not intended to replace medical diagnosis.

Keywords: *Iron Deficiency Anemia, Random Forest, Machine Learning, Classification System, Model Evaluation.*