

ABSTRAK

Stunting pada balita merupakan masalah kesehatan global yang berdampak jangka panjang pada perkembangan fisik dan kognitif. Penelitian ini bertujuan menerapkan kinerja tiga algoritma *machine learning*, yaitu *Random Forest*, *LightGBM*, dan *XGBoost*, dalam memprediksi *stunting* pada balita berdasarkan data status gizi. Dataset yang digunakan mencakup 16.546 catatan pengukuran balita dari Puskesmas Tasikmalaya (Januari-Oktober 2024), dengan fitur utama seperti berat badan, tinggi badan, lingkaran lengan atas (LiLA), umur, jenis kelamin, dan cara pengukuran. Tahap *preprocessing* meliputi data *cleaning*, transformasi, seleksi fitur menggunakan *information gain*, normalisasi *min-max*, dan *balancing* data dengan SMOTE. Pembagian data menggunakan *5-fold*, *7-fold*, dan *10-fold cross-validation*, serta pengujian dengan variasi jumlah atribut (1-10) dan parameter *n_estimator* (30, 50, 100, dan 200).

Hasil penelitian menunjukkan bahwa *XGBoost* mencapai akurasi tertinggi 97.85% pada pengujian dengan 10 fitur (termasuk ekstraksi fitur perubahan berat, tinggi, LiLA, dan jumlah pengukuran), menggunakan *10-fold* dan *n_estimator* = 200. Fitur paling berpengaruh adalah Berat, Tinggi, LiLA, Umur_Bulan, Cara Ukur, dan JK, sementara ekstraksi fitur meningkatkan akurasi secara signifikan. Proses *balancing* data dengan SMOTE terbukti efektif meningkatkan akurasi model pada dataset tidak seimbang.

Dalam penelitian ini algoritma *XGBoost* merupakan algoritma terbaik untuk prediksi *stunting* pada balita. Model ini dapat diimplementasikan dalam sistem GUI untuk membantu tenaga medis dan masyarakat dalam identifikasi dini *stunting* secara efisien. Saran pengembangan meliputi eksplorasi teknik ekstraksi fitur tambahan dan optimasi hiperparameter lebih lanjut.

Kata Kunci : *Stunting*, *Random Forest*, *LightGBM*, *XGBoost*, SMOTE, Klasifikasi, Prediksi.

ABSTRACT

Stunting in children under five is a global health problem that has a long-term impact on physical and cognitive development. This study aims to applying the performance of three machine learning algorithms, namely Random Forest, LightGBM, and XGBoost, in predicting stunting in toddlers based on nutritional status data. The dataset used includes 16,546 measurement records of under-fives from Tasikmalaya Community Health Center (January-October 2024), with key features such as weight, height, upper arm circumference (LiLA), age, gender, and measurement method. The preprocessing stage includes data cleaning, transformation, feature selection using information gain, min-max normalization, and data balancing with SMOTE. Data sharing uses 5-fold, 7-fold, and 10-fold cross-validation, and testing with variations in the number of attributes (1-10) and $n_estimator$ parameters (30, 50, 100, and 200).

The results showed that XGBoost achieved the highest accuracy of 97.85% in the test with 10 features (including the features of weight change extraction, height, LiLA, and number of measurements), using 10-fold and $n_estimator = 200$. The most influential features were height, weight, and LiLA, while feature extraction improved accuracy significantly. The data balancing process with SMOTE proved effective in improving model accuracy on unbalanced datasets.

In this study, the XGBoost algorithm is the best algorithm for stunting prediction in toddlers. This model can be implemented in a GUI system to help medical personnel and the public in early identification of stunting efficiently. Development suggestions include exploring additional feature extraction techniques and further hyperparameter optimization.

Keywords: *Stunting, Random Forest, LightGBM, XGBoost, SMOTE, Classification, Prediction.*