

ABSTRAK

Status gizi balita merupakan indikator penting dalam pemantauan kesehatan dan pertumbuhan anak. Namun, pengolahan data status gizi sering menghadapi permasalahan ketidakseimbangan kelas yang dapat menurunkan kinerja model klasifikasi, terutama pada kategori gizi kurang dan gizi buruk. Penelitian ini bertujuan untuk merancang dan mengimplementasikan model klasifikasi status gizi balita berbasis data antropometri menggunakan algoritma Random Forest yang dioptimalkan, dengan mengacu pada standar Z-score *World Health Organization* (WHO) dan ketentuan Peraturan Menteri Kesehatan (Permenkes). Dataset yang digunakan terdiri dari 2035 data balita, meliputi 1043 balita laki-laki dan 992 balita perempuan, dengan total 14 variabel sebelum *preprocessing*. Setelah proses penentuan label menggunakan pendekatan *majority voting*, distribusi data terdiri dari 855 data gizi buruk, 817 data gizi normal, 217 data gizi over, dan 138 data gizi kurang, yang menunjukkan ketidakseimbangan antar kelas. Indikator antropometri yang digunakan mencakup BB/U, TB/U, BB/TB, serta atribut pendukung lainnya. Untuk meningkatkan kualitas data, dilakukan *outlier removal* menggunakan metode *Interquartile Range* (IQR) pada kelas gizi normal serta penanganan ketidakseimbangan data menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE). Model Random Forest dibangun dengan pendekatan *majority voting* dan dievaluasi menggunakan metrik *akurasi*, *precision*, *recall*, dan *F1-score*. Hasil pengujian menunjukkan akurasi sebesar 83% dengan F1-score macro 0,82, serta peningkatan kinerja yang signifikan pada kelas minoritas, khususnya gizi kurang dan gizi over, dibandingkan skenario tanpa penanganan ketidakseimbangan data. Dengan demikian, kombinasi Random Forest, outlier removal, dan SMOTE mampu memperbaiki hasil klasifikasi status gizi balita, terutama pada kelas dengan jumlah data yang lebih sedikit.

Kata Kunci : Data Antropometri, *Outlier Removal*, *Random Forest*, Status Gizi Balita , *Synthetic Minority Over-sampling Technique* (SMOTE)

ABSTRACT

Nutritional status of toddlers is an important indicator for monitoring child health and growth. However, the processing of nutritional status data often encounters class imbalance problems, which can reduce the performance of classification models, particularly in the undernutrition and severe malnutrition categories. This study aims to design and implement a toddler nutritional status classification model based on anthropometric data using an optimized Random Forest algorithm, in accordance with the World Health Organization (WHO) Z-score standards and the regulations of the Ministry of Health. The dataset used consists of 2,035 toddler records, comprising 1,043 male and 992 female toddlers, with a total of 14 variables prior to preprocessing. After label determination using a majority voting approach, the data distribution includes 855 severe malnutrition samples, 817 normal nutrition samples, 217 overnutrition samples, and 138 undernutrition samples, indicating a significant class imbalance. The anthropometric indicators employed include weight-for-age (WFA), height-for-age (HFA), weight-for-height (WFH), along with additional supporting attributes. To improve data quality, outlier removal using the Interquartile Range (IQR) method was applied to the normal nutrition class, and class imbalance was addressed using the Synthetic Minority Over-sampling Technique (SMOTE). The Random Forest model was constructed using a majority voting approach and evaluated using accuracy, precision, recall, and F1-score metrics. Experimental results demonstrate an accuracy of 83% and a macro-averaged F1-score of 0.82, with a significant performance improvement observed in minority classes, particularly undernutrition and overnutrition, compared to scenarios without imbalance handling. Therefore, the combination of Random Forest, outlier removal, and SMOTE is proven to be effective in enhancing the reliability of toddler nutritional status classification.

Keywords: Toddlers, Status Gizi Balita, Random Forest, Data Antropometri, , Outlier Removal, , Synthetic Minority Over-sampling Technique (SMOTE)