

ABSTRAK

Stunting masih menjadi masalah kesehatan serius di Indonesia yang memengaruhi tumbuh kembang anak. Penelitian ini bertujuan membandingkan kinerja dua algoritma klasifikasi, yaitu K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM), dalam mengidentifikasi status stunting pada balita berdasarkan data antropometri. Dataset berjumlah 16.500 data diambil dari Kaggle dengan tujuh atribut utama, seperti jenis kelamin, usia, berat dan panjang lahir, berat dan panjang badan, serta status ASI. Data diproses melalui tahapan pembersihan, transformasi data kategorik, dan penyeimbangan menggunakan undersampling dan oversampling. Model diuji menggunakan k-fold cross-validation dan tuning hyperparameter melalui Grid Search dan Random Search. Hasil eksperimen menunjukkan bahwa akurasi tertinggi untuk KNN diperoleh dengan Grid Search pada data oversampling sebesar 88,81%. Sedangkan SVM dengan Grid Search menghasilkan akurasi, yakni 83,28% pada data oversampling. Namun, akurasi tertinggi secara keseluruhan dicapai oleh algoritma SVM dengan Random Search menggunakan kernel sigmoid dan linear, yaitu sebesar 89,18% pada data oversampling dan 89,99% pada data undersampling dengan kernel polynomial. Temuan ini menunjukkan bahwa pemilihan metode tuning dan kernel sangat berpengaruh terhadap performa model. Dengan demikian, SVM dengan Random Search terbukti lebih unggul dibandingkan KNN dalam klasifikasi stunting. Penelitian ini diharapkan dapat mendukung pengembangan sistem deteksi dini stunting yang lebih akurat untuk membantu upaya pencegahan di bidang kesehatan anak.

Kata kunci: Stunting, Klasifikasi, K-Nearest Neighbor, Support Vector Machine, Hyperparameter Tuning, Grid Search, Random Search.

ABSTRACT

Stunting remains a critical public health issue in Indonesia, affecting the growth and development of children. This study aims to compare the performance of two classification algorithms, K-Nearest Neighbor (KNN) and Support Vector Machine (SVM), in identifying stunting status in toddlers based on anthropometric data. The dataset, obtained from Kaggle, consists of 16,500 records with seven key attributes, including gender, age, birth weight and length, body weight and length, and breastfeeding status. The data processing includes cleaning, categorical transformation, and balancing using undersampling and oversampling techniques. Model evaluation was conducted using k-fold cross-validation and hyperparameter tuning through Grid Search and Random Search. Experimental results show that the highest accuracy for KNN was achieved using Grid Search on oversampled data, reaching 83.28%. Similarly, SVM with Grid Search on oversampled data also achieved an accuracy of 83.28%. However, the highest overall performance was obtained by SVM with Random Search, which reached 89.18% accuracy on oversampled data (using sigmoid and linear kernels) and 89.99% accuracy on undersampled data with a polynomial kernel. These findings indicate that kernel selection and tuning methods significantly affect model performance. In conclusion, SVM with Random Search outperformed KNN in classifying stunting status. This research is expected to contribute to the development of more accurate early detection systems for stunting, thereby supporting preventive efforts in child health management.

Keywords: Stunting, Classification, K-Nearest Neighbor, Support Vector Machine, Hyperparameter Tuning, Grid Search, Random Search.