

ABSTRAK

Stunting merupakan permasalahan gizi kronis yang berdampak serius terhadap tumbuh kembang anak. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Naïve Bayes* dan *K-Nearest Neighbors* (KNN) dalam memprediksi stunting pada balita di Kecamatan Banjaran. Dataset terdiri dari 12.000 data balita dengan tiga fitur utama: usia, jenis kelamin, dan tinggi badan. Penelitian menggunakan pendekatan kuantitatif dengan penerapan algoritma pembelajaran mesin. Teknik oversampling SMOTE diterapkan hanya pada data latih untuk menghindari kebocoran data, serta digunakan validasi silang 5-fold cross-validation. Nilai $K=3$ dipilih untuk model akhir KNN berdasarkan analisis kurva validasi guna menghindari overfitting. Hasil menunjukkan bahwa KNN secara signifikan mengungguli *Naïve Bayes* dalam semua metrik evaluasi. Model *Naïve Bayes* menghasilkan akurasi 67,50%, presisi 50,87%, recall 61,38%, F1 score 55,63%, spesifisitas 70,54%, dan AUC score 75,71%. Sementara itu, model KNN ($K=3$) mencapai akurasi 99,11%, presisi 98,08%, recall 99,25%, F1 score 98,66%, spesifisitas 99,03%, dan AUC score 99,65%. Perbedaan kinerja kedua model dikonfirmasi melalui Uji McNemar dengan nilai $p < 0,05$, yang menunjukkan perbedaan signifikan secara statistik. Kinerja rendah *Naïve Bayes* disebabkan oleh pelanggaran asumsi independensi fitur, khususnya korelasi tinggi antara usia dan tinggi badan ($r \approx 0.87$). Kesimpulannya, KNN merupakan algoritma yang lebih tepat digunakan untuk prediksi stunting pada dataset ini. Namun, keterbatasan jumlah fitur menunjukkan perlunya penelitian lanjutan dengan variabel tambahan dan validasi eksternal sebelum diterapkan dalam skala yang lebih luas.

Kata Kunci: Stunting, *Machine Learning*, Klasifikasi, *Naïve Bayes*, *K-Nearest Neighbors* (KNN), Pra-pemrosesan Data, SMOTE

ABSTRACT

Stunting is a chronic nutritional problem that seriously impacts child growth and development. This study aims to compare the performance of the Naïve Bayes and K-Nearest Neighbors (KNN) algorithms in predicting stunting in toddlers in Banjaran District. The dataset consists of 10,060 toddler data points with three main features: age, gender, and height. The research employed a quantitative approach by applying machine learning algorithms. The SMOTE oversampling technique was applied only to the training data to avoid data leakage, and 5-fold cross-validation was used. A K-value of 3 was selected for the final KNN model based on validation curve analysis to prevent overfitting. The results show that KNN significantly outperformed Naïve Bayes across all evaluation metrics. The Naïve Bayes model yielded an accuracy of 67.50%, precision of 50.87%, recall of 61.38%, F1-score of 55.63%, specificity of 70.54%, and an AUC score of 75.71%. Meanwhile, the KNN (K=3) model achieved an accuracy of 99.11%, precision of 98.08%, recall of 99.25%, F1-score of 98.66%, specificity of 99.03%, and an AUC score of 99.65%. The performance difference between the two models was confirmed by McNemar's Test with a p -value < 0.05 , indicating a statistically significant difference. The low performance of Naïve Bayes was attributed to the violation of the feature independence assumption, particularly the high correlation between age and height ($r \approx 0.87$). In conclusion, KNN is the more appropriate algorithm for stunting prediction on this dataset. However, the limitation of features suggests the need for further research with additional variables and external validation before wider-scale implementation.

Keywords: *Stunting, Machine Learning, Classification, Naïve Bayes, K-Nearest Neighbors (KNN), Data Pre-processing, SMOTE*