

ABSTRAK

Kerawanan pangan rumah tangga di Kota Langsa, Aceh, masih memerlukan sistem deteksi yang akurat untuk mendukung perumusan kebijakan intervensi yang tepat sasaran. Keterbatasan sumber daya dan ketimpangan akses pangan antarwilayah menuntut pendekatan berbasis data yang mampu mengidentifikasi rumah tangga rentan secara efisien dan objektif. Tujuan penelitian ini adalah menerapkan dan membandingkan performa algoritma *Double Random Forest* (DRF) dengan *Random Forest* (RF) konvensional serta RF yang dikombinasikan dengan *EasyEnsemble* dalam memprediksi status kerawanan pangan rumah tangga. Data bersumber dari Survei Sosial Ekonomi Nasional (SUSENAS) tahun 2023 untuk wilayah Kota Langsa, mencakup 2.057 rumah tangga dengan 13 variabel prediktor sosial-ekonomi, sementara variabel target berupa label biner yang dibentuk dari tiga indikator berat *Food Insecurity Experience Scale* (FIES), menghasilkan distribusi 36,7 persen rumah tangga rawan pangan dan 63,3 persen tidak rawan. Ketiga model dilatih dan diuji pada 618 data uji, kemudian stabilitasnya diverifikasi melalui 50 kali repetisi dengan uji ANOVA satu arah. Hasil pengujian menunjukkan performa yang relatif setara antar ketiga model, dengan DRF mencatat AUC tertinggi sebesar 0,8758, diikuti RF+*EasyEnsemble* sebesar 0,8711 dan RF *baseline* sebesar 0,8696, sementara RF+*EasyEnsemble* unggul pada sensitivitas (76,65 persen) dan akurasi (79,61 persen). Uji ANOVA mengonfirmasi bahwa selisih performa antar model tidak signifikan secara statistik ($F=1,9263$; $p=0,1493$). Analisis *Feature Importance* menunjukkan status penerima bantuan sosial rutin sebagai variabel paling dominan, berkontribusi 41,05 persen terhadap klasifikasi. Temuan ini menegaskan bahwa akses terhadap program perlindungan sosial menjadi determinan utama kerawanan pangan rumah tangga, dan pemilihan algoritma sebaiknya mempertimbangkan kebutuhan praktis di lapangan alih-alih semata keunggulan statistik.

Kata kunci: *Double Random Forest*, kerawanan pangan, SUSENAS, *feature importance*, klasifikasi data tidak seimbang

ABSTRACT

Household food insecurity in Langsa City, Aceh, continues to require an accurate detection system to support well-targeted policy interventions. Limited resources and uneven food access across regions demand a data-driven approach capable of identifying vulnerable households efficiently and objectively. This study aimed to implement and compare the performance of the Double Random Forest (DRF) algorithm against conventional Random Forest (RF) and RF combined with EasyEnsemble in predicting household food insecurity status. Data were obtained from the 2023 National Socioeconomic Survey (SUSENAS) for Langsa City, covering 2,057 households with 13 socioeconomic predictor variables, while the target variable was a binary label built from three severe indicators of the Food Insecurity Experience Scale (FIES), yielding 36.7 percent food-insecure and 63.3 percent food-secure households. The three models were trained and tested on 618 samples, with stability verified through 50 repeated trials using a one-way ANOVA test. Results showed comparable performance across models, with DRF recording the highest AUC (0.8758), followed by RF+EasyEnsemble (0.8711) and baseline RF (0.8696), while RF+EasyEnsemble outperformed the others in sensitivity (76.65 percent) and accuracy (79.61 percent). The ANOVA test confirmed that performance differences among models were not statistically significant ($F=1.9263$; $p=0.1493$). Feature importance analysis identified routine social assistance recipient status as the most dominant variable, contributing 41.05 percent to the classification outcome. These findings indicate that access to social protection programs is the primary determinant of household food insecurity, and algorithm selection should therefore prioritize practical field requirements rather than statistical superiority alone.

Keywords: Double Random Forest; food insecurity; SUSENAS; feature importance; imbalanced data classification