

INTISARI

Readmisi rumah sakit dalam 30 hari merupakan indikator penting kualitas layanan, khususnya pada pasien obstetri dan ginekologi (*obgyn*) yang berisiko komplikasi pasca perawatan. Penelitian ini membandingkan performa *Logistic Regression* (*LR*) dan *Extreme Gradient Boosting* (*XGBoost*) dalam memprediksi readmisi pasien *obgyn* menggunakan data rekam medis elektronik 109.079 pasien rawat inap (2021–2023) dengan 19 atribut klinis. Dataset yang tidak seimbang diatasi dengan *Synthetic Minority Oversampling Technique* (*SMOTE*). Evaluasi model dilakukan menggunakan akurasi, AUC, F1-Score, presisi, dan recall; tambahan optimasi dilakukan melalui *threshold tuning* (0,32) pada *LR* serta teknik *stacking*. Hasil menunjukkan *LR* baseline mencapai *F1-Score* = 84,23%, *AUC* = 70,30%, yang meningkat menjadi *F1-Score* = 86,45%, *AUC* = 87,74% setelah *threshold tuning*, dengan *recall* = 92,37%. *XGBoost* memberikan performa terbaik *F1-Score* = 98,11%, *AUC* = 97,01%, sedangkan *stacking* menghasilkan performa sebanding. Analisis *Odds Ratio* dari *LR* mengindikasikan variabel seperti lama rawat ($OR \approx 0,29$) menurunkan peluang readmisi, sedangkan spesialis ($OR \approx 1,42$) meningkatkan peluang. Analisis *SHAP* pada *XGBoost* menegaskan bahwa lama rawat inap, selisih waktu kunjungan, dan riwayat penyakit merupakan faktor dominan. Hasil penelitian ini menyimpulkan bahwa *XGBoost* lebih unggul dalam akurasi prediksi readmisi, sedangkan *LR*—meski kinerjanya meningkat melalui *threshold tuning*—tetap lebih rendah. Namun, *LR* memiliki kelebihan interpretasi klinis melalui *OR*, sehingga keduanya memberikan manfaat komplementer: *XGBoost* untuk prediksi operasional, dan *LR* untuk pemahaman faktor risiko.

Kata kunci: readmisi, pasien *obgyn*, logistic regression, *XGBoost*, *SHAP*, *stacking* model

ABSTRACT

Hospital readmission within 30 days is an important indicator of service quality, especially for obstetrics and gynecology (obgyn) patients who are at risk of complications after treatment. This study compares the performance of Logistic Regression (LR) and Extreme Gradient Boosting (XGBoost) in predicting obgyn patient readmissions using electronic medical records data from 109,079 inpatients (2021–2023) with 19 clinical attributes. The imbalanced dataset was addressed using the Synthetic Minority Oversampling Technique (SMOTE). Model evaluation was performed using accuracy, AUC, F1-Score, precision, and recall; additional optimization was performed through threshold tuning (0.32) on LR and stacking techniques. The results showed that the baseline LR achieved an F1-Score = 84.23%, AUC = 70.30%, which increased to F1-Score = 86.45%, AUC = 87.74% after threshold tuning, with recall = 92.37%. XGBoost provided the best performance with an F1-Score = 98.11% and AUC = 97.01%, while stacking produced comparable performance. Odds Ratio analysis from LR indicated that variables such as length of stay ($OR \approx 0.29$) decreased the probability of readmission, while specialists ($OR \approx 1.42$) increased the probability. SHAP analysis on XGBoost confirms that length of stay, time difference between visits, and medical history are dominant factors. The results of this study conclude that XGBoost is superior in readmission prediction accuracy, while LR—despite its improved performance through threshold tuning—remains inferior. However, LR has the advantage of clinical interpretation through OR, so both provide complementary benefits: XGBoost for operational prediction and LR for understanding risk factors.

Keywords: readmission, obgyn patients, logistic regression, XGBoost, SHAP, stacking model