

ABSTRAK

Tuberkulosis (TB) tetap menjadi tantangan kesehatan global yang kritis, di mana Indonesia menempati peringkat beban kasus tertinggi kedua di dunia. Penelitian ini bertujuan untuk mengoptimalkan klasifikasi penyakit TB menggunakan algoritma *Extreme Gradient Boosting (XGBoost)* melalui pendekatan *hyperparameter tuning* untuk mengatasi keterbatasan diagnosis konvensional yang cenderung lambat. Menggunakan 7.150 data rekam medis yang bersumber dari RS Syarifah Ambami Rato Ebuh, Puskesmas, dan Kaggle, metodologi penelitian mencakup *preprocessing*, penanganan data tidak seimbang dengan *Synthetic Minority Over-sampling Technique (SMOTE)*, serta *Grid Search* yang dipadukan dengan *10-Fold Cross-Validation*. Hasil penelitian menunjukkan bahwa model *XGBoost* dengan tuning secara konsisten mengungguli model tanpa tuning di seluruh rasio pengujian, dengan akurasi tertinggi mencapai 84,71% pada rasio data 70:30. Analisis perbandingan performa yang memprioritaskan metrik *Recall* dan minimalisasi *False Negative (FN)* mengungkapkan bahwa model tanpa tuning cenderung memiliki bias terhadap kelas mayoritas serta mengalami gejala *overfitting* ringan. Sebaliknya, model yang menggunakan tuning terbukti lebih stabil, bahkan berhasil mencapai angka nol kasus *False Positive* pada rasio 90:10. Model final ini diimplementasikan ke dalam aplikasi berbasis website menggunakan *framework Flask*.

Kata kunci: *Grid Search, Hyperparameter Tuning, SMOTE, Tuberkulosis, XGBoost.*

ABSTRACT

Tuberculosis (TB) remains a critical global health challenge, with Indonesia ranking second for the highest disease burden in the world. This study aims to optimize TB disease classification using the Extreme Gradient Boosting (XGBoost) algorithm through a hyperparameter tuning approach to address the limitations of conventional diagnosis, which tends to be slow. Utilizing 7,150 medical record datasets sourced from Syarifah Ambami Rato Ebuh Hospital, Public Health Centers (Puskesmas), and Kaggle, the research methodology encompasses preprocessing, addressing data imbalance using the Synthetic Minority Over-sampling Technique (SMOTE), and Grid Search integrated with 10-Fold Cross-Validation. The results indicate that the tuned XGBoost model consistently outperforms the non-tuned model across all test ratios, with the highest accuracy of 84.71% achieved at a 70:30 data split ratio. Performance comparison analysis, prioritizing the Recall metric and the minimization of False Negatives (FN), reveals that the non-tuned model tends to exhibit bias toward the majority class and suffers from mild overfitting. In contrast, the tuned model proved to be more stable, even achieving zero False Positive cases at a 90:10 ratio. This final model was implemented into a web-based application using the Flask framework.

Keywords: *Grid Search, Hyperparameter Tuning, SMOTE, Tuberculosis, XGBoost.*