

ABSTRAK

Tuberkulosis (TB) paru merupakan salah satu penyakit menular yang masih menjadi permasalahan kesehatan sehingga diperlukan upaya deteksi dan klasifikasi yang efektif. Penelitian ini bertujuan untuk mengklasifikasikan penyakit TB paru menggunakan algoritma *Classification and Regression Tree* (CART) dan *Extreme Gradient Boosting* (XGBOOST) serta membandingkan kinerja kedua algoritma tersebut. Data yang digunakan berupa 3.732 data rekam medis pasien dari Rumah Sakit Khusus Paru Sumatera Utara dengan atribut umur, jenis kelamin, demam, batuk lebih dari 2 minggu, berat badan turun, keringat malam, sesak napas, nafsu makan turun, riwayat kontak TB, dan diabetes mellitus. Tahapan penelitian meliputi *preprocessing*, transformasi data, pembagian data *training* dan *testing* dengan rasio 80:20, pembentukan model CART dan XGBOOST, serta evaluasi menggunakan *Accuracy*, *Precision*, *Recall*, *F1-Score*, *Confusion Matrix*, dan *10-Fold Cross Validation*. Hasil penelitian menunjukkan bahwa CART memperoleh *Accuracy* sebesar 69,86%, *Precision* sebesar 72,36%, *Recall* sebesar 58,94%, dan *F1-Score* sebesar 64,96%. Sementara itu, XGBOOST memperoleh *Accuracy* sebesar 70,64%, *Precision* sebesar 74,47%, *Recall* sebesar 57,95%, dan *F1-Score* sebesar 65,18%. Berdasarkan hasil evaluasi, XGBOOST dipilih sebagai model terbaik karena memperoleh *Accuracy*, *Precision*, dan *F1-Score* yang lebih tinggi dibandingkan CART, meskipun nilai *Recall* CART sedikit lebih tinggi. Model XGBOOST selanjutnya diimplementasikan ke dalam sistem berbasis web untuk membantu proses klasifikasi awal TB paru berdasarkan gejala pasien. Penelitian ini menunjukkan bahwa CART dan XGBOOST dapat diterapkan sebagai pendekatan klasifikasi TB paru berbasis data rekam medis.

Kata kunci: Tuberkulosis Paru, Klasifikasi, CART, XGBOOST, *Machine Learning*.

ABSTRACT

Pulmonary tuberculosis (TB) is an infectious disease that remains a public health problem, requiring effective detection and classification efforts. This study aims to classify pulmonary TB using the Classification and Regression Tree (CART) and Extreme Gradient Boosting (XGBOOST) algorithms and compare the performance of both algorithms. The data used consisted of 3,732 patient medical records obtained from the North Sumatra Special Lung Hospital, with attributes including age, sex, fever, cough for more than two weeks, weight loss, night sweats, shortness of breath, decreased appetite, history of TB contact, and diabetes mellitus. The research stages included preprocessing, data transformation, splitting the data into training and testing sets with an 80:20 ratio, development of CART and XGBOOST models, and evaluation using Accuracy, Precision, Recall, F1-Score, Confusion Matrix, and 10-Fold Cross Validation. The results showed that CART achieved an Accuracy of 69.86%, Precision of 72.36%, Recall of 58.94%, and F1-Score of 64.96%. Meanwhile, XGBOOST achieved an Accuracy of 70.64%, Precision of 74.47%, Recall of 57.95%, and F1-Score of 65.18%. Based on the evaluation results, XGBOOST was selected as the best model because it achieved higher Accuracy, Precision, and F1-Score than CART, although CART achieved a slightly higher Recall. The XGBOOST model was subsequently implemented into a web-based system to assist in the initial classification of pulmonary TB based on patient symptoms. This study demonstrates that CART and XGBOOST can be applied as Machine Learning approaches for pulmonary TB classification using medical record data.

Keywords: *Pulmonary Tuberculosis, Classification, CART, XGBOOST, Machine Learning.*