

## INTISARI

### **PENANGANAN DATA TIDAK SEIMBANG MENGGUNAKAN SMOTE PADA KASUS KLASIFIKASI TUBERKULOSIS BERBASIS CITRA *CHEST X-RAY***

Oleh:

Muhammad Fadhlullah Kh.TQ

22/500709/PPA/06361

Penelitian ini mendalami permasalahan ketidakseimbangan dataset dalam klasifikasi citra *Chest X-Ray* (CXR) TBX11K dengan menerapkan metode Random Forest (RF) dan XGBoost (XGB) dengan atau tanpa teknik *resampling Synthetic Minority Over-sampling Technique* (SMOTE). Tujuan penelitian ini adalah mengukur pengaruh SMOTE terhadap performa model dalam klasifikasi citra CXR TBX11K.

Dalam penelitian ini, teknik SMOTE diterapkan pada model klasifikasi RF dan XGB. Penggunaan SMOTE bertujuan untuk meningkatkan jumlah sampel kelas minoritas (TB positif) agar tidak terlalu senjang terhadap jumlah kelas mayoritas (TB negatif). Setiap model diukur dengan metrik evaluasi yang sama agar dapat dibandingkan, seperti akurasi, presisi, *recall*, dan skor F1.

Hasil penelitian menunjukkan bahwa penggunaan teknik SMOTE pada model RF dan XGB efektif dalam mengatasi ketidakseimbangan kelas pada dataset. Model RF tanpa SMOTE mencapai akurasi sekitar 93,33%, sementara model RF dengan SMOTE mencapai akurasi 92,72%. Model XGB tanpa SMOTE mencapai akurasi 94,11%, sedangkan XGB dengan SMOTE mencapai akurasi 94,33%. Meskipun SMOTE meningkatkan performa model, terdapat kendala dalam memprediksi kelas minoritas 'altb' dan 'ltb', yang sulit diatasi bahkan dengan *resampling*, disebabkan fitur yang kurang representatif. Dari hasil eksperimen, model XGB dengan SMOTE merupakan model yang paling optimal untuk klasifikasi citra TBX11K.

**Kata kunci:** Tuberkulosis, Random Forest, XGBoost, *Chest X-Ray*, *machine learning*, *imbalance data*, SMOTE, VGG16

## **ABSTRACT**

### ***HANDLING IMBALANCED DATA USING SMOTE IN TUBERCULOSIS CLASSIFICATION BASED ON CHEST X-RAY IMAGES***

By:

Muhammad Fadhlullah Kh.TQ

22/500709/PPA/06361

*This research delves into the issue of dataset imbalance in the classification of Chest X-Ray (CXR) images in TBX11K by applying the Random Forest (RF) and XGBoost (XGB) methods with or without the Synthetic Minority Over-sampling Technique (SMOTE) resampling technique. The objective of this study is to assess the impact of SMOTE on model performance in the classification of CXR TBX11K images.*

*In this research, the SMOTE technique is applied to the RF and XGB classification models. The use of SMOTE aims to increase the number of minority class samples (TB positive) to mitigate the imbalance with the majority class samples (TB negative). Each model is evaluated using the same metrics for comparison, such as accuracy, precision, recall, and F1 score.*

*The research results indicate that the use of the SMOTE technique on RF and XGB models is effective in addressing class imbalance in the dataset. The RF model without SMOTE achieves an accuracy of approximately 93.33%, while the RF model with SMOTE achieves an accuracy of 92.72%. The XGB model without SMOTE achieves an accuracy of 94.11%, while the XGB model with SMOTE achieves an accuracy of 94.33%. Although SMOTE improves the model's performance, there are challenges in predicting minority classes 'altb' and 'ltb,' which are difficult to overcome even with resampling due to less representative features. From the experimental results, the XGB model with SMOTE is the most optimal model for classifying TBX11K images.*

**Keywords:** *Tuberculosis, Random Forest, XGBoost, Chest X-Ray, machine learning, imbalance data, SMOTE, VGG16*