ABSTRAK

Rozai, Muhammad, 2026, **KLASIFIKASI AKTIVITAS MANUSIA MENGGUNAKAN ALGORITMA *RANDOM FOREST* PADA *IMBALANCED SMARTPHONE SENSOR* DATASET.** Tugas Akhir, Program Sarjana, Program Studi Teknik Informatika, Universitas Muhammadiyah Jember.

Pembimbing: (1) Agung Nilogiri, ST., M.Kom.; (2) Taufiq Timur W., S.Kom., M.Kom.

Human Activity Recognition (HAR) berbasis sensor smartphone memiliki potensi untuk digunakan dalam berbagai bidang, seperti pemantauan aktivitas dan keamanan. Namun, ketidakseimbangan kelas pada dataset dapat memengaruhi kinerja model klasifikasi, terutama dalam mengenali kelas minoritas. Penelitian ini bertujuan menganalisis kinerja algoritma *Random Forest* sebelum dan sesudah *balancing* data serta menganalisis pengaruh *hyperparameter tuning* terhadap kinerja model dalam mengklasifikasikan aktivitas *Standing* dan *Walking*. Dataset yang digunakan adalah *Accelerometer Gyro Mobile Phone* dari UCI Machine Learning Repository yang terdiri atas 31.991 data. Setelah proses *cleaning*, diperoleh 31.847 data dengan rasio ketidakseimbangan kelas sekitar 1:54,8 antara kelas *Standing* dan *Walking*. Tahapan penelitian meliputi *preprocessing*, normalisasi menggunakan MinMaxScaler, penanganan ketidakseimbangan data latih menggunakan SMOTE dengan *k-neighbors* = 5, serta pengujian empat kondisi pemodelan, yaitu *Random Forest default* dan *Random Forest* dengan *hyperparameter tuning*, masing-masing pada data *imbalanced* dan data hasil SMOTE. Pengujian dilakukan menggunakan skenario *stratified split* 70:30 dan 80:20. Evaluasi menggunakan *Accuracy*, *Macro Precision*, *Macro Recall*, dan *Macro F1-Score*, dengan *Macro F1-Score* sebagai metrik utama. Hasil pengujian menunjukkan bahwa penerapan SMOTE meningkatkan *Macro F1-Score* pada kedua skenario pembagian data. Pada split 70:30, nilai *Macro F1-Score* meningkat dari 63,48% menjadi 69,12%, sedangkan pada split 80:20 meningkat dari 61,44% menjadi 67,88%. Sebaliknya, pengaruh *hyperparameter tuning* tidak menunjukkan peningkatan yang konsisten. Pada data *imbalanced*, *Macro F1-Score* setelah *tuning* sebesar 61,91% pada split 70:30 dan 61,44% pada split 80:20. Pada data hasil SMOTE, nilainya sebesar 69,12% pada split 70:30 dan 68,26% pada split 80:20. Dengan demikian, *hyperparameter tuning* tidak meningkatkan *Macro F1-Score* pada split 70:30, sedangkan peningkatan sebesar 0,38 poin persentase pada split 80:20 relatif kecil dibandingkan variasi antar-*seed*. *Macro F1-Score* tertinggi (69,12%) diperoleh pada data hasil SMOTE split 70:30, baik tanpa maupun dengan *hyperparameter tuning*, karena parameter terbaik hasil *GridSearchCV* identik dengan parameter default. Hasil robustness check menggunakan sepuluh nilai *random seed* pada Kondisi D menunjukkan simpangan baku sebesar 0,64 poin persentase pada split 70:30 (rentang 67,52%–69,47%) dan 1,19 poin persentase pada split 80:20 (rentang 67,14%–70,61%). Simpangan baku pada split 80:20 lebih besar daripada peningkatan akibat tuning, sehingga pengaruh *hyperparameter tuning* pada penelitian ini tidak konsisten dan tidak dapat dinyatakan berarti.

Kata kunci: *random forest*, SMOTE, *class imbalance*, *human activity recognition*, *hyperparameter tuning*.

ABSTRACT

Rozai, Muhammad, 2026, ***HUMAN ACTIVITY CLASSIFICATION USING THE RANDOM FOREST ALGORITHM ON THE IMBALANCED SMARTPHONE SENSOR DATASET.*** *Final Project, Undergraduate Program, Informatics Engineering Study Program, University of Muhammadiyah Jember.*

Supervisors: (1) Agung Nilogiri, ST., M.Kom.; (2) Taufiq Timur W., S.Kom., M.Kom.

Human Activity Recognition (HAR) based on smartphone sensors has the potential to be used in various fields, such as activity monitoring and security. However, class imbalance in a dataset can affect the performance of classification models, particularly in recognizing the minority class. This study aims to analyze the performance of the Random Forest algorithm before and after data balancing, and to analyze the effect of hyperparameter tuning on model performance in classifying Standing and Walking activities. The dataset used is the Accelerometer Gyro Mobile Phone dataset from the UCI Machine Learning Repository, which consists of 31,991 records. After cleaning, 31,847 records remained, with a class imbalance ratio of approximately 1:54.8 between the Standing and Walking classes. The research stages include preprocessing, normalization using MinMaxScaler, handling of training data imbalance using SMOTE with k-neighbors = 5, and testing of four modeling conditions, namely default Random Forest and Random Forest with hyperparameter tuning, each applied to imbalanced data and SMOTE-balanced data. Testing was conducted using stratified split scenarios of 70:30 and 80:20. Evaluation used Accuracy, Macro Precision, Macro Recall, and Macro F1-Score, with Macro F1-Score as the primary metric. The results show that applying SMOTE increased the Macro F1-Score in both data split scenarios. In the 70:30 split, the Macro F1-Score increased from 63.48% to 69.12%, while in the 80:20 split it increased from 61.44% to 67.88%. In contrast, the effect of hyperparameter tuning did not show a consistent improvement. On imbalanced data, the Macro F1-Score after tuning was 61.91% in the 70:30 split and 61.44% in the 80:20 split. On SMOTE-balanced data, it was 69.12% in the 70:30 split and 68.26% in the 80:20 split. Accordingly, hyperparameter tuning did not improve the Macro F1-Score in the 70:30 split, while the 0.38 percentage-point gain in the 80:20 split was small relative to the variation across random seeds. The highest Macro F1-Score (69.12%) was obtained on SMOTE-balanced data in the 70:30 split, both without and with hyperparameter tuning, because the best parameters selected by GridSearchCV were identical to the default parameters. A robustness check using ten random seeds on Condition D showed standard deviations of 0.64 percentage points in the 70:30 split (range 67.52%–69.47%) and 1.19 percentage points in the 80:20 split (range 67.14%–70.61%). The standard deviation in the 80:20 split is larger than the gain from tuning, so the effect of hyperparameter tuning in this study is inconsistent and cannot be considered meaningful.

Keywords: *random forest, SMOTE, class imbalance, human activity recognition, hyperparameter tuning.*