BAB 1
PENDAHULUAN

1.1 Latar Belakang

Pesatnya kemajuan teknologi *smartphone* saat ini menghadirkan perangkat yang dilengkapi berbagai sensor, seperti akselerometer dan giroskop, yang mampu merekam data pergerakan manusia secara *real-time*. Sensor yang terdapat pada *smartphone* dapat dimanfaatkan untuk menangkap pola gerak tubuh tanpa memerlukan perangkat tambahan. Pemanfaatan data sensor tersebut untuk mengidentifikasi aktivitas manusia dikenal sebagai *Human Activity Recognition* (HAR). HAR berbasis sensor *smartphone* dapat diterapkan dalam berbagai bidang, seperti pemantauan kesehatan, keamanan, olahraga, dan sistem rumah pintar. Hal ini didukung oleh penelitian Munid Alanazi (Alanazi dkk., 2024) yang menunjukkan bahwa penggunaan sensor inersial *smartphone* dengan pendekatan *machine learning* mampu mengklasifikasikan aktivitas manusia dengan akurasi yang tinggi.

Dalam penerapannya, sistem HAR menggunakan berbagai algoritma *machine learning* untuk mengklasifikasikan aktivitas berdasarkan data sensor. Beberapa algoritma yang telah digunakan antara lain *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Naive Bayes*, dan metode *ensemble* seperti *Random Forest*.

Random Forest merupakan salah satu algoritma ensemble berbasis bagging yang diperkenalkan oleh Leo Breiman (Breiman, 2001). Algoritma ini membangun beberapa pohon keputusan dan menggabungkan keluaran setiap pohon, yang pada implementasi scikit-learn dilakukan dengan merata-ratakan probabilitas kelas dari seluruh pohon. Random Forest memiliki beberapa kelebihan, seperti mampu menangani data dengan banyak fitur, lebih tahan terhadap overfitting dibandingkan satu decision tree, serta dapat digunakan pada data yang memiliki noise (Thakur & Biswas, 2023). Namun, Random Forest juga memiliki kekurangan. Proses pelatihannya membutuhkan waktu lebih lama ketika jumlah pohon (n_estimators) semakin banyak. Selain itu, pada data dengan distribusi kelas yang tidak seimbang, model dapat lebih banyak mempelajari pola dari kelas mayoritas sehingga kelas minoritas kurang terdeteksi (Chawla dkk., 2002; Widiawati & Yunita, 2025). Oleh

karena itu, random forest digunakan dalam penelitian ini untuk mengklasifikasikan aktivitas *standing* dan *walking* berdasarkan data *sensor smartphone*, dengan SMOTE dan *hyperparameter tuning* sebagai upaya meningkatkan kinerjanya.

Dataset yang digunakan dalam penelitian ini adalah Accelerometer Gyro Mobile Phone Dataset yang dikembangkan oleh Abdullah AlSahly dan diakses melalui UCI Machine Learning Repository (AlSahly dkk., 2022). Dataset tersebut berisi data sensor *accelerometer* dan *gyroscope* pada *smartphone* untuk merepresentasikan aktivitas manusia. Penelitian ini menggunakan dua kelas aktivitas, yaitu *Standing* dan *Walking*. Berdasarkan hasil pembacaan dataset, diperoleh 31.991 data awal. Setelah dilakukan tahap *cleaning* dengan menghapus 144 data duplikat, jumlah data yang digunakan dalam penelitian menjadi 31.847 data. Kelas *Standing* berjumlah 571 data atau 1,79%, sedangkan kelas *Walking* berjumlah 31.276 data atau 98,21%. Distribusi tersebut menunjukkan adanya ketidakseimbangan kelas (*class imbalance*) yang sangat ekstrem, dengan perbandingan kelas sekitar 1:54,8. Kondisi ini dapat menyebabkan model lebih banyak mempelajari pola dari kelas mayoritas sehingga kemampuan dalam mengenali kelas *Standing* sebagai kelas minoritas menjadi kurang optimal.

Untuk menangani ketidakseimbangan tersebut, penelitian ini menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE) yang diperkenalkan oleh Nitesh V. Chawla (Chawla dkk., 2002). SMOTE membentuk data sintetis baru pada kelas minoritas berdasarkan data yang sudah ada. Dengan cara tersebut, jumlah data kelas minoritas pada data latih dapat ditingkatkan sehingga distribusi kelas menjadi lebih seimbang. Penelitian Chyntia Raras Ajeng Widiawati & Ika Romadoni Yunita (Widiawati & Yunita, 2025) menunjukkan bahwa penerapan SMOTE dapat meningkatkan kinerja klasifikasi pada data sensor yang tidak seimbang. Dalam penelitian ini, SMOTE diterapkan pada data latih dengan nilai *k-neighbors* sebesar 5, sedangkan data uji tidak dilakukan SMOTE.

Selain penanganan ketidakseimbangan kelas, penelitian ini menggunakan *hyperparameter tuning* untuk mencari konfigurasi *Random Forest* yang sesuai dengan data penelitian. Proses tersebut dilakukan menggunakan GridSearchCV dengan *5-fold cross-validation* dan *F1-Score Macro* sebagai kriteria pemilihan parameter terbaik. Parameter yang diuji meliputi *n_estimators*, *max_depth*, dan

max_features. Proses pencarian parameter dilakukan pada data tidak seimbang maupun data yang telah diseimbangkan menggunakan SMOTE.

Berdasarkan tahapan tersebut, penelitian ini membandingkan empat kondisi pemodelan, yaitu *Random Forest default* pada data tidak seimbang, *Random Forest* dengan GridSearchCV pada data tidak seimbang, *Random Forest default* pada data yang telah diseimbangkan menggunakan SMOTE, serta *Random Forest* dengan GridSearchCV pada data yang telah diseimbangkan menggunakan SMOTE. Masing-masing kondisi diuji menggunakan *Stratified Train-Test Split* dengan pembagian data 70:30 dan 80:20. Kinerja model dinilai menggunakan *Accuracy*, *Precision Macro*, *Recall Macro*, dan *F1-Score Macro*, dengan *F1-Score Macro* sebagai metrik utama. Pengujian tersebut dilakukan untuk mengetahui pengaruh SMOTE dan *hyperparameter tuning* terhadap kinerja *Random Forest* pada data sensor *smartphone* dengan ketidakseimbangan kelas yang ekstrem.

1.2 Rumusan Masalah

1. Berapa tingkat kinerja algoritma *Random Forest* sebelum *hyperparameter tuning* dalam mengklasifikasikan aktivitas manusia (berdiri dan berjalan) berdasarkan nilai *Accuracy, Precision, Recall*, dan *F1-Score* pada skenario pembagian data 70:30 dan 80:20?
2. Berapa tingkat kinerja algoritma *Random Forest* setelah dilakukan *balancing* data (penanganan *class imbalance*) dalam mengklasifikasikan aktivitas manusia pada skenario pembagian data 70:30 dan 80:20?
3. Bagaimana pengaruh *hyperparameter tuning* terhadap kinerja algoritma *Random Forest* dalam mengklasifikasikan aktivitas manusia sebelum dan sesudah penanganan *class imbalance?*

1.3 Tujuan Penelitian

1. Mengetahui tingkat kinerja algoritma Random Forest sebelum *hyperparameter tuning* dalam mengklasifikasikan aktivitas manusia pada skenario pembagian data 70:30 dan 80:20.
2. Mengetahui tingkat kinerja algoritma Random Forest setelah penanganan *class imbalance* (balancing data) pada skenario pembagian data 70:30 dan 80:20.
3. Menganalisis pengaruh *hyperparameter tuning* terhadap kinerja algoritma Random Forest dalam mengklasifikasikan aktivitas manusia sebelum dan sesudah penanganan *class imbalance*.

1.4 Batasan Penelitian

1. Dataset: Penelitian ini menggunakan *Accelerometer Gyro Mobile Phone Dataset* yang diambil dari *UCI Machine Learning Repository* (AlSahly dkk., 2022). Dataset mentah terdiri dari 31.991 sampel sebelum melalui tahap *preprocessing*.
2. Klasifikasi Aktivitas: Ruang lingkup aktivitas manusia dibatasi pada klasifikasi biner (*binary classification*), yaitu aktivitas *Standing* (berdiri) dan *Walking* (berjalan).
3. Fitur: Fitur yang digunakan dalam pemodelan mencakup enam data mentah sensor, yaitu *accX*, *accY*, *accZ* (akselerometer) dan *gyroX*, *gyroY*, *gyroZ* (giroskop).
4. Algoritma: Algoritma *machine learning* yang digunakan adalah *Random Forest* dengan *hyperparameter tuning* menggunakan *GridSearchCV*.
5. Penanganan Imbalanced Data: Ketidakseimbangan kelas diatasi dengan teknik *SMOTE* (*Synthetic Minority Oversampling Technique*) yang diterapkan khusus pada *data latih* (*training data*) guna menghindari kebocoran data (*data leakage*).
6. Pembagian Data: Skenario *data splitting* yang diterapkan adalah 70:30 dan 80:20. Metode *stratified split* digunakan untuk menjaga konsistensi proporsi distribusi kelas antara *data latih* dan *data uji* (*testing data*).

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi dalam bidang informatika, khususnya pada penerapan *machine learning* untuk klasifikasi aktivitas manusia (*Human Activity Recognition*) menggunakan algoritma *Random Forest* pada dataset *smartphone* dengan kondisi ketidakseimbangan kelas yang ekstrem (*extreme class imbalance*). Secara akademis, penelitian ini dapat menambah referensi mengenai kinerja algoritma *Random Forest* dengan penerapan *hyperparameter tuning* serta teknik SMOTE dalam menangani ketidakseimbangan kelas pada dataset. Secara praktis, hasil penelitian ini dapat memberikan gambaran mengenai kinerja *Random Forest* pada beberapa kondisi pemodelan, termasuk penerapan *hyperparameter tuning* dan SMOTE, serta dapat menjadi bahan pertimbangan dalam penelitian selanjutnya dengan karakteristik dataset yang serupa.