ABSTRAK

Lokalisasi *indoor* berbasis *Bluetooth Low Energy* (BLE) menghadapi fluktuasi *Received Signal Strength Indicator* (RSSI), sinyal yang *sparse*, dan ketidakseimbangan kelas. Penelitian ini menerapkan *imbalance-aware stacking ensemble* untuk mengklasifikasikan 21 lokasi tingkat ruangan pada *nursing care facility* menggunakan *dataset* ABC 2026. Sebanyak 1.673.395 observasi BLE dari 25 *beacon* dikelompokkan dalam *temporal window* 5 detik sehingga menghasilkan 9.012 *fingerprint* dengan 100 fitur statistik dan tingkat sparsitas 86,04%. *Temporal alignment* dengan label menghasilkan 601 *fingerprint* berlabel dengan *imbalance ratio* 51,33:1. Data kemudian diimputasi, distandardisasi, dan diseleksi menjadi 50 fitur menggunakan *Mutual Information*. SMOTE-Tomek pada data latih awal menurunkan *imbalance ratio* dari 61,50:1 menjadi 1,23:1. *Logistic Regression*, *Random Forest*, LightGBM, dan CatBoost digabungkan melalui *stacking* dengan *Logistic Regression* sebagai *meta-learner*, sedangkan *soft voting* digunakan sebagai pembanding. Pada data *validation* sebanyak 506 sampel, *stacking ensemble* memperoleh *accuracy* 95,06% dan *Macro F1-score* 94,76%, tertinggi di antara enam model, tetapi hanya unggul satu prediksi benar dibandingkan CatBoost (94,86% dan 94,58%). *F1-score* per kelas berkisar dari 0,7317 pada *cafeteria* sampai 1,0000 pada lima ruang bernomor. Karena data *validation* dibentuk dari data hasil *resampling* dan data *holdout* belum dievaluasi, hasil ini menunjukkan bahwa *stacking ensemble* kompetitif terhadap model individual terbaik, sedangkan kemampuan generalisasinya pada data asli masih perlu diuji.

Kata Kunci: *BLE RSSI; indoor localization; class imbalance; SMOTE-Tomek; stacking ensemble*

ABSTRACT

Bluetooth Low Energy (BLE)-based indoor localization faces fluctuating Received Signal Strength Indicator (RSSI) values, sparse signals, and class imbalance. This study applies an imbalance-aware stacking ensemble to classify 21 room-level locations in a nursing care facility using the ABC 2026 dataset. 1,673,395 BLE observations from 25 beacons were grouped into 5-second temporal windows, producing 9,012 fingerprints with 100 statistical features and 86.04% sparsity. Temporal alignment yielded 601 labeled fingerprints with an imbalance ratio of 51.33:1. Data were imputed, standardized, and reduced to 50 features using Mutual Information. SMOTE-Tomek on the initial training data reduced the imbalance ratio from 61.50:1 to 1.23:1. Logistic Regression, Random Forest, LightGBM, and CatBoost were combined through stacking with a Logistic Regression meta-learner, with soft voting as a comparison. On 506 validation samples, the stacking ensemble achieved 95.06% accuracy and a 94.76% Macro F1-score, the highest among six models, but outperformed CatBoost (94.86% and 94.58%) by only one correct prediction. Per-class F1-scores ranged from 0.7317 for the cafeteria to 1.0000 for five numbered rooms. Because validation data came from resampled data and the holdout set was not evaluated, the stacking ensemble is competitive with the best individual model, while generalization to original data remains untested.

Keywords: *BLE RSSI; indoor localization; class imbalance; SMOTE-Tomek; stacking ensemble*