ABSTRAK

Pengobatan presisi untuk kanker payudara memerlukan model prediktif yang akurat guna mengatasi heterogenitas respons pasien terhadap terapi standar. Salah satu pendekatan potensial untuk menangkap variabilitas tersebut adalah penggunaan profil molekuler. Data proteomik telah terbukti memiliki nilai prediktif yang tinggi, namun penentuan arsitektur *machine learning* yang paling optimal untuk memanfaatkannya masih menjadi tantangan. Penelitian ini bertujuan untuk mengoptimasi arsitektur model prediksi respons obat dengan membandingkan efektivitas metode representasi fitur obat klasik dan modern. Evaluasi komparatif dilakukan terhadap sepuluh arsitektur model yang merupakan kombinasi dari dua metode ekstraksi fitur (*Morgan Fingerprint* dan *Graph Neural Network*) dengan lima algoritma *machine learning* (LightGBM, XGBoost, *Random Forest*, SVM, dan MLP). Dataset terdiri dari 47 *cell line* kanker payudara dan 239 obat yang divalidasi menggunakan metode *Group Shuffle Split* (80:20) untuk mencegah kebocoran data. Kinerja model diukur menggunakan metrik utama *Area Under Precision-Recall Curve* (AUPRC). Hasil eksperimen menunjukkan bahwa algoritma berbasis *boosting* (LightGBM dan XGBoost) secara konsisten mengungguli algoritma lainnya. Representasi fitur berbasis graf (GNN) terbukti memberikan informasi topologi yang lebih informatif dibandingkan *Morgan Fingerprint* pada model berbasis pohon. Arsitektur model paling optimal yang ditemukan adalah kombinasi fitur GNN dengan algoritma XGBoost, yang mencapai skor AUPRC tertinggi sebesar 0,9844. Temuan tersebut menegaskan bahwa integrasi representasi molekuler berbasis graf dengan algoritma XGBoost merupakan strategi yang efektif untuk meningkatkan akurasi pemodelan farmakogenomik pada data proteomik.

Kata Kunci: Prediksi Respons Obat, Kanker Payudara, Proteomik, Farmakogenomik, *Group Shuffle Split*, Machine Learning, *Graph Neural Network.*