

ABSTRAK

IMPLEMENTASI PENGGABUNGAN EKSTRAKSI FITUR DALAM MENINGKATKAN KLASIFIKASI *POST-TRANSLATIONAL MODIFICATION* (PTM) GLIKOSILASI PADA PROTEIN *SEQUENCE N, O DAN C* DENGAN METODE *EXTREME GRADIENT BOOSTING* (XGBOOST)

Oleh

DAMAYANTI

Post-translational modification (PTM) adalah salah satu mekanisme penting dalam mengatur fungsi protein. Modifikasi pasca translasi mengacu pada penambahan modifikasi protein kovalen dan enzimatik dalam biosintesis protein, yang memiliki peran penting dalam memodifikasi fungsi protein dan mengatur ekspresi gen. Salah satu modifikasi pasca translasi adalah glikosilasi. Glikosilasi adalah penambahan gugus gula ke struktur protein. Glikosilasi telah terkait dengan beberapa penyakit diantaranya diabetes, kanker, dan alzheimer. Oleh karena itu, penting untuk mengantisipasi terjadinya glikosilasi dengan melakukan prediksi glikosilasi.

Permasalahan dalam prediksi glikosilasi saat ini masih bergantung pada teknik laboratorium manual, yang menyebabkan proses prediksi menjadi lambat dan memerlukan biaya peralatan laboratorium yang tinggi. Untuk mengatasi hal tersebut, diperlukan pendekatan *machine learning* sehingga prediksi dapat dilakukan lebih cepat dan tidak membutuhkan biaya yang mahal.

Data yang digunakan dalam penelitian ini adalah data PTM glikosilasi-N, glikosilasi-O, dan glikosilasi-C yang diperoleh dari website UniProt yang tersedia secara terbuka. Penelitian ini bertujuan untuk meningkatkan akurasi klasifikasi modifikasi pasca translasi glikosilasi-N, glikosilasi-O, dan glikosilasi-C dengan menggabungkan 5 (lima) Ekstraksi fitur dan menggunakan algoritma Extreme Gradient Boosting (XGBoost).

Ekstraksi fitur terdiri dari: AAIndex, *Hydrophobicity*, SABLE, CTD, dan PseAAC. Seleksi fitur dilakukan dengan pendekatan MRMR. Masing-masing fitur memberikan kontribusi terhadap peningkatan prediksi glikosilasi. Fitur AAIndex memberikan kontribusi terbesar pada peningkatan prediksi glikosilasi-N secara keseluruhan sebesar 24%. Sedangkan, fitur SABLE memberikan kontribusi terbesar pada peningkatan

prediksi glikosilasi-O sebesar 44%. Fitur *Hydrophobicity* dan PseAAC masing-masing berkontribusi sebesar 27% untuk peningkatan prediksi glikosilasi-C

Hasil penelitian ini menunjukkan kinerja prediksi modifikasi pasca translasi glikosilasi-N, glikosilasi-O dan glikosilasi-C dengan masing-masing nilai akurasi 100%. Pendekatan menggunakan XGBoost dalam penelitian ini berhasil meningkatkan akurasi sebesar 5% dibandingkan dengan penelitian sebelumnya.

Kata Kunci: *Post-translational modifications*, glikosilasi, *sequence*, xgboost, klasifikasi