

BAB 1 PENDAHULUAN

1.1. Latar Belakang

Osteoporosis adalah gangguan tulang yang ditandai dengan rendahnya massa tulang dan kerusakan mikroarsitektur jaringan tulang yang berakibat pada peningkatan kerapuhan tulang dan risiko patah tulang. Lebih dari 200 juta orang terkena dampaknya di seluruh dunia, dan osteoporosis merupakan penyakit tulang yang paling banyak ditemukan [1]. Secara klinis, penyakit ini sering kali tidak terdiagnosa hingga terjadi patah tulang akibat osteoporosis. Risiko osteoporosis seumur hidup pada wanita dan pria masing-masing berkisar antara 40%-50% dan 13%-22%, dan pasien yang menderitanya menunjukkan gangguan kualitas hidup yang berhubungan dengan kesehatan [1]. Untuk mengatasi masalah ini, diperlukan target molekuler yang efektif dalam memperlambat degradasi tulang, salah satunya adalah enzim Cathepsin K (CatK). CatK merupakan enzim yang terkenal dalam proses perombakan sel-sel tulang yang sudah tua dan menjadi target penting dalam pengembangan terapi osteoporosis [2].

Dalam dekade terakhir, upaya penelitian ditujukan pada pengembangan inhibitor CatK untuk pengobatan osteoporosis dan gangguan tulang lainnya yang terkait dengan keropos tulang patologis. Bukti baru yang muncul menunjukkan bahwa selain resorpsi tulang, Cathepsin K terlibat dalam pergantian protein matriks ekstraseluler di organ, seperti paru-paru, tiroid dan kulit, dan memainkan peran penting dalam penyakit kardiovaskular, peradangan dan obesitas [3]. Oleh karena itu, pengembangan inhibitor CatK telah menjadi fokus penelitian dalam menemukan solusi farmakologis yang efektif untuk mengatasi penyakit yang melibatkan degradasi tulang berlebihan.

Strategi penemuan obat konvensional digunakan untuk mengembangkan obat baru, tetapi proses dari identifikasi hingga uji klinis membutuhkan waktu lebih dari 12 tahun [4]. Untuk meningkatkan efisiensi dalam memprediksi bioaktivitas inhibitor CatK, pendekatan *machine learning* telah menarik minat yang cukup besar karena potensinya untuk mempercepat penemuan obat dalam hal waktu, tenaga kerja, dan biaya [4]. Berbagai penelitian telah dilakukan terkait penerapan *machine learning* dalam memprediksi bioaktivitas inhibitor untuk pengobatan penyakit tulang, terutama yang berkaitan dengan inhibitor CatK.

Penelitian oleh Ilyas et al. (2024) menunjukkan bahwa kombinasi *Quantitative Structure-Activity Relationship* (QSAR), *Molecular docking*, dan simulasi *Molecular Dynamics* (MD) dengan algoritma *machine learning Extreme Trees* (ET) dapat secara efektif memprediksi bioaktivitas inhibitor CatK, dengan akurasi 0.977 [5]. Parwez et al. (2024) juga melakukan penelitian untuk memprediksi bioaktivitas inhibitor CatK dengan model *Support Vector Machine* (SVM) dan mendapatkan akurasi 0.997 [6]. Dibia et al. (2022) melakukan penelitian menggunakan pendekatan QSAR untuk memprediksi bioaktivitas inhibitor Cyclooxygenase-2 (COX-2) dengan model XGBoost dan mendapatkan akurasi 0.9484 [7]. Penelitian oleh Zhang et al. (2023) menunjukkan bahwa model RFE-XGBoost mencapai akurasi sebesar 97.56% dalam prediksi penyakit tulang belakang [8]. Chakravarti (2020) melakukan penelitian untuk reseptor protein Aryl Hydrocarbon Receptor (AHR) menggunakan metode *Logistic regression* berbasis ECFP (Extended Connectivity Fingerprint), dengan akurasi 83.48% [9].

Beberapa penelitian yang sudah dilakukan, secara umum hasilnya sudah cukup baik, tetapi masih terdapat ruang untuk perbaikan. Salah satu alternatif untuk perbaikan adalah melalui seleksi fitur yang lebih optimal. Algoritma metaheuristik, seperti *Glowworm Swarm Optimization* (GSO), telah menunjukkan kemampuan untuk mencari solusi global secara efisien, menjadikannya alat yang efektif untuk seleksi fitur dan optimasi

hyperparameter. GSO memiliki keunggulan dalam eksplorasi ruang pencarian yang kompleks dan adaptif terhadap masalah non-linear [10]. Algoritma GSO dipilih karena keunggulannya dalam menangani masalah optimisasi multi-modal. Tidak seperti algoritma *swarm intelligence* lain yang cenderung konvergen pada satu titik optimal tunggal, GSO mampu mengidentifikasi beberapa subset fitur potensial secara bersamaan melalui mekanisme jangkauan dinamisnya, dimana bisa terdapat lebih dari satu kombinasi fitur yang optimal. Model prediksi berbasis XGBoost memiliki potensi besar karena kemampuannya dalam menangani dataset berukuran besar dan kompleks [11]. XGBoost dipilih karena rekam jejaknya yang terbukti unggul dalam menangani data tabular, terutama dalam domain medis dan bioinformatika. Keunggulan XGBoost tidak hanya terletak pada akurasinya yang tinggi, tetapi juga pada efisiensi komputasinya yang didukung oleh pemrosesan paralel dan kemampuannya untuk mencegah *overfitting* melalui mekanisme regularisasi. Dengan mengintegrasikan GSO dan XGBoost, diharapkan dapat diperoleh model prediksi yang lebih akurat dan efisien untuk memprediksi bioaktivitas inhibitor.

Penelitian ini bertujuan untuk mengembangkan model prediksi bioaktivitas inhibitor Cathepsin K untuk pengobatan osteoporosis dengan menggunakan kombinasi algoritma *Glowworm Swarm Optimization* (GSO) dan XGBoost. GSO digunakan untuk seleksi fitur dan dapat membantu mencari solusi global optimum yang sering sulit dicapai dengan metode konvensional. XGBoost diterapkan untuk membangun model prediksi yang mampu menangani dataset kompleks dan besar.

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah yang dapat diambil adalah:

1. Bagaimana efektivitas *Glowworm Swarm Optimization* (GSO) dalam seleksi fitur untuk prediksi bioaktivitas inhibitor Cathepsin K?

2. Bagaimana pengaruh optimasi hyperparameter dengan GSO terhadap performa model XGBoost?
3. Apakah kombinasi GSO dan XGBoost mampu memberikan performa yang optimal dalam prediksi bioaktivitas inhibitor Cathepsin K?

1.3. Tujuan dan Manfaat

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi sebuah model prediksi bioaktivitas untuk senyawa inhibitor Cathepsin K (CatK). Secara spesifik, tujuan yang hendak dicapai adalah mengevaluasi efektivitas algoritma Glowworm Swarm Optimization (GSO) dalam melakukan seleksi fitur dari data senyawa kimia, menganalisis pengaruh dari optimasi hyperparameter terhadap kinerja model klasifikasi XGBoost, serta menguji apakah kombinasi GSO dan XGBoost mampu menghasilkan performa prediksi yang optimal. Pencapaian tujuan-tujuan tersebut diukur secara kuantitatif melalui metrik evaluasi performa, yang mencakup Accuracy, Precision, Recall, dan F1-Score.

Manfaat utama dari penelitian ini adalah menyediakan sebuah model komputasi yang diharapkan dapat meningkatkan efisiensi dan produktivitas dalam tahap awal penemuan obat untuk penyakit tulang seperti osteoporosis. Dengan adanya model yang mampu memprediksi bioaktivitas secara cepat dan akurat, proses penyaringan senyawa kimia potensial dapat dipercepat secara signifikan. Hal ini dapat mengurangi waktu dan biaya yang besar, yang menjadi tantangan utama dalam metode penemuan obat konvensional, sehingga para peneliti dapat lebih fokus pada kandidat senyawa yang paling menjanjikan.

1.4. Batasan Masalah

Untuk memastikan penelitian dapat berjalan secara fokus dan layak untuk dikerjakan dalam batasan waktu dan sumber daya yang tersedia, maka ditetapkan beberapa batasan masalah sebagai berikut:

1. Fitur yang digunakan untuk merepresentasikan struktur kimia senyawa dalam penelitian ini terbatas pada molekul 3D, khususnya yang berasal dari database PubChem. Penelitian ini tidak menggunakan atau mengombinasikan jenis deskriptor lain seperti deskriptor *fingerprints*.
2. Proses seleksi fitur dalam penelitian ini hanya menggunakan algoritma metaheuristik *Glowworm Swarm Optimization* (GSO). Penelitian ini tidak melakukan studi komparatif dengan algoritma metaheuristik atau metode seleksi fitur lainnya. Pembatasan ini bertujuan untuk dapat mengevaluasi secara mendalam efektivitas dari GSO secara spesifik untuk kasus prediksi bioaktivitas inhibitor CatK.
3. Model *machine learning* yang dikembangkan dan dioptimalkan untuk klasifikasi terbatas pada algoritma *Extreme Gradient Boosting* (XGBoost). Kinerja dari fitur-fitur yang dihasilkan oleh GSO tidak diujikan pada model klasifikasi lain seperti *Support Vector Machine* (SVM) atau *Deep Learning*, agar penelitian lebih fokus pada optimasi interaksi antara GSO dan XGBoost.
4. Sumber dan Lingkup Dataset. Dataset yang digunakan dalam penelitian ini bersumber dari satu basis data, yaitu ChEMBL, dengan ID 268 yang spesifik untuk inhibitor Cathepsin K. Model yang dibangun tidak diuji pada dataset dari sumber lain atau untuk target protein yang berbeda. Ini dilakukan untuk menjaga konsistensi dan fokus analisis pada target terapi yang telah ditentukan.

1.5. Metode Penelitian

Penyelesaian Tugas Akhir ini dilakukan dengan menggunakan pendekatan yang terstruktur dan sistematis. Penelitian diawali dengan studi literatur secara mendalam untuk mengumpulkan informasi dari berbagai sumber relevan seperti jurnal dan artikel ilmiah. Tahap ini bertujuan untuk membangun landasan teori yang kuat mengenai algoritma *Glowworm Swarm*

Optimization (GSO) dan *Extreme Gradient Boosting* (XGBoost), serta memahami penelitian terkini terkait prediksi bioaktivitas inhibitor Cathepsin K. Setelah landasan teori terbentuk, penelitian dilanjutkan dengan tahap perancangan sistem, yang mencakup pembuatan alur kerja penelitian dalam bentuk diagram alir (*flowchart*). Proses perancangan ini merinci setiap langkah, mulai dari preparasi data, alur seleksi fitur menggunakan GSO, hingga skema optimasi dan validasi model XGBoost.

Tahap berikutnya adalah implementasi, di mana rancangan sistem diwujudkan ke dalam kode program. Pada tahap ini, dilakukan pengolahan dataset, penerapan algoritma GSO untuk seleksi fitur, serta pelatihan model XGBoost dengan fitur-fitur yang telah terpilih. Model yang telah diimplementasikan kemudian dievaluasi kinerjanya. Hasil dari pengukuran ini kemudian melalui proses analisis statistik untuk menilai keandalan model. Analisis ini dilakukan dengan menggunakan metrik evaluasi yang dihitung dari confusion matrix, seperti Accuracy dan F1-Score, untuk memastikan model mampu memberikan prediksi yang akurat dan dapat dipertanggungjawabkan.

1.6. Jadwal Pelaksanaan

Tabel 1.6. Jadwal Pelaksanaan Tugas Akhir.

No.	Deskripsi Tahapan	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5	Bulan 6
1	Studi Literatur						
2	Pengumpulan Data						
3	Perancangan Model						
4	Implementasi Model						
5	Pengujian dan Analisis						
6	Penyusunan Buku TA						