

Penerapan Algoritma Firefly-Ensemble untuk Memprediksi Toksisitas: Studi Kasus Toksisitas AR-LBD

Andini Aprilia Putri¹, Isman Kurniawan²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹andiniprli@student.telkomuniversity.ac.id, ²ismankrn@telkomuniversity.ac.id

Abstrak

Penelitian ini bertujuan untuk mengimplementasikan algoritma Firefly-Ensemble dalam memprediksi toksisitas dengan studi kasus *Androgen Receptor Ligand Binding Domain* (AR-LBD). Toksisitas merupakan faktor utama penyebab kegagalan obat dalam uji klinis, dan prediksi toksisitas yang akurat sangat penting untuk mengoptimalkan proses pengembangan obat. Penelitian ini menggunakan Firefly Algorithm untuk seleksi fitur dan metode Ensemble, yang mencakup Random Forest, AdaBoost, dan XGBoost, untuk pengoptimalan parameter model. Hasil penelitian menunjukkan bahwa penggunaan Firefly Algorithm sebagai seleksi fitur dapat meningkatkan akurasi model pada beberapa model, terutama XGBoost, dengan akurasi sebesar 99,30%, F1-score 99,31%, recall 99,88%, dan precision 98,75%. Selain itu, penerapan pengoptimalan parameter menggunakan metode Ensemble juga berkontribusi pada peningkatan kinerja model, dengan XGBoost menunjukkan performa terbaik, diikuti oleh Random Forest dan AdaBoost. Hasil ini menunjukkan bahwa kombinasi algoritma Firefly-Ensemble meningkatkan akurasi prediksi toksisitas, sehingga dapat digunakan untuk mendukung pengembangan metode *in silico* dalam toksikologi dan penemuan obat.

Kata kunci : Pembelajaran Mesin, Algoritma Firefly, Ensemble, AR-LBD, Toksisitas

Abstract

This research aims to implement the Firefly-Ensemble algorithm in predicting the toxicity with the case study of the Androgen Receptor Ligand Binding Domain (AR-LBD). Toxicity is the main factor causing drug failure in clinical trials, and accurate toxicity prediction is crucial for optimizing the drug development process. This study uses the Firefly Algorithm for feature selection and Ensemble methods, which include Random Forest, AdaBoost, and XGBoost, for model parameter optimization. The research results show that the use of the Firefly Algorithm for feature selection can improve the model's accuracy on several models, especially XGBoost, with an accuracy of 99.30%, an F1-score of 99.31%, a recall of 99.88%, and a precision of 98.75%. Additionally, the application of parameter optimization using the Ensemble method also contributes to the improvement of model performance, with XGBoost showing the best performance, followed by Random Forest and AdaBoost. These results indicate that the combination of the Firefly-Ensemble algorithm improves the accuracy of toxicity prediction, thus it can be used to support the development of in silico methods in toxicology and drug discovery.

Keywords: Machine Learning, Firefly Algorithm, Ensemble, AR-LBD, Toxicity

1. Pendahuluan

Latar Belakang

Obat adalah suatu objek yang dibuat oleh para tenaga medis untuk membantu mencegah, menyembuhkan, dan meningkatkan kesehatan pasien [1]. Toksisitas adalah sejauh mana suatu zat dapat membahayakan suatu organisme hidup. Toksisitas dalam pengembangan obat mengacu pada efek berbahaya atau reaksi buruk yang dapat disebabkan oleh obat dalam tubuh. Ini bisa termasuk efek samping, kerusakan organ, reaksi alergi, atau efek buruk lainnya pada kesehatan. Penilaian dan manajemen toksisitas sangat penting dalam pengembangan obat untuk memastikan bahwa obat itu aman dan efektif untuk pasien [2]. Salah satu bagian penting dari proses penemuan obat adalah pengukuran toksisitas. Ini membantu mengidentifikasi dan memprioritaskan senyawa yang memiliki potensi terbesar untuk digunakan secara aman dan efektif pada manusia dan juga mengurangi risiko kegagalan yang mahal pada proses tahap akhir penemuan obat [3], hal ini telah dilaporkan menjadi penyebab sekitar 30% kandidat obat ditarik [4].

Empat faktor yang dapat menyebabkan 90% kegagalan klinis dalam pengembangan obat, menurut analisis data uji klinis dari tahun 2010–2017: kurangnya efikasi klinis (40–50%), toksisitas yang tidak dapat dikelola (30%), sifat obat yang buruk (10–15%) dan kurangnya kebutuhan komersial dan perencanaan strategis yang buruk (10–15%) [2]. Kegagalan obat disebabkan oleh toksisitas yang sangat tinggi sehingga tidak dapat diterima. Ini menyumbang sekitar satu per lima kegagalan uji klinis dan dua per tiga penarikan obat di seluruh dunia

setelah diluncurkan [5]. Oleh karena itu, tugas toksikologi adalah mengidentifikasi risiko tertentu untuk kasus tertentu yang dapat menyebabkan dampak negatif yang disebabkan oleh interaksi antara dosis dan paparan toksik, mekanisme toksik, dan elemen yang mempengaruhi kasus tertentu [6].

Terdapat beberapa metode konvensional telah digunakan dalam memprediksi profil toksisitas bahan kimia, biasanya teknik *in vitro* dan *in vivo* digunakan untuk memeriksa keamanan obat, termasuk efek samping dan toksisitas. Namun metode *in vitro* dan *in vivo* sering dibatasi oleh etika, waktu, anggaran, dan sumber daya lainnya. Oleh karena itu, metode alternatif untuk prediksi toksisitas *in silico* diperlukan karena sangat penting untuk pengambilan keputusan regulasi dan pemilihan dalam pembuatan obat [7],[8]. Pada tahap awal pengembangan obat, pengembangan model *in silico* untuk menyaring dan memahami mekanisme toksisitas obat mungkin sangat membantu karena ini dapat mengurangi biaya dan waktu tenaga kerja [4]. Profil toksisitas senyawa kimia dapat diprediksi dengan banyak teknik komputasi. Tinjauan ini memberikan penjelasan menyeluruh tentang penerapan algoritma pembelajaran mesin dalam toksikologi prediktif [7]. Proses pembelajaran mesin adalah salah satu pendekatan yang efektif untuk digunakan dalam bidang biomedis. Ini ditunjukkan oleh bagaimana mesin pembelajaran mendekati pendekatan pembuatan algoritma otomatis yang berkualitas tinggi yang dapat digunakan untuk diagnosis atau prediksi penyakit [9].

Penelitian tentang prediksi toksisitas telah dilakukan menggunakan berbagai metode pembelajaran mesin. Pada tahun 2021, Feng, et al. melakukan penelitian untuk memprediksi toksisitas reproduksi zat kimia menggunakan metode Ensemble dan sidik jari molekuler, dengan menggunakan model Ensemble yang dikembangkan menggunakan metode Support Vector Machine, Random Forest, dan Extreme Gradient Boosting, serta sembilan sidik jari molekuler yang dihitung untuk kumpulan data yang berisi 1823 zat kimia. Model Ensemble-Top12 memiliki kinerja prediksi terbaik, dengan akurasi (ACC) sebesar 86,33% [10]. Penelitian yang dilakukan oleh Riwandy, et al. pada tahun 2024 bertujuan untuk memanfaatkan kekuatan *Artificial Neural Networks* (ANNs) yang dioptimalkan oleh *Cuckoo Search Algorithm* (CSA) dan menggunakan dataset sidik jari untuk memprediksi toksisitas. Hasilnya, dengan skor F1 0,6153 dan akurasi 0,9652, menunjukkan kinerja yang baik, ini menunjukkan kemampuan untuk memprediksi toksisitas sudah baik [11]. Penelitian yang dilakukan oleh Aditya et al. pada tahun 2023 bertujuan untuk menggunakan metode ANN yang dioptimalkan oleh algoritma *grey wolf* untuk memprediksi toksisitas senyawa kimia dengan sidik jari berbasis fitur. Hasil tes data Tox21 Data Challenge menunjukkan skor F1 dan akurasi data tes masing-masing 0,590 dan 0,887 [12].

Dara et al. pada tahun 2021 pernah melakukan tinjauan tentang penggunaan pembelajaran mesin dalam penemuan obat. Dalam kasus ini, sebuah profil untuk efektivitas ekspresi gen biomarker, profil efektivitas ini digunakan untuk analisis pengobatan farmakologis dengan akurasi 88,9% untuk Classification Tree dan 83,3% untuk model Random Forest [13]. Pada tahun 2021, Zhang, et al. melakukan penelitian tentang pengembangan dan evaluasi model prediksi *in silico* untuk toksisitas pernapasan yang diinduksi obat dengan menggunakan metode klasifikasi Naive Bayes. Dalam penelitian ini, model prediksi Naive Bayes untuk toksisitas pernapasan telah dibuat menggunakan kumpulan data 1241 senyawa. Klasifier Naive Bayes menghasilkan akurasi prediksi keseluruhan untuk set pelatihan sebesar 91,8% dan untuk set uji eksternal sebesar 84,3% [14]. Meskipun hasil penelitian sebelumnya cukup baik, masih ada ruang untuk peningkatan. Penelitian sebelumnya kurang optimal dalam hal seleksi fitur dan pengoptimalan parameter yang mempengaruhi performa prediksi. Penelitian ini bertujuan untuk mengisi celah tersebut dengan mengimplementasikan algoritma Firefly-Ensemble yang lebih efisien untuk seleksi fitur dan pengoptimalan parameter. Oleh karena itu, seleksi fitur sangat penting karena digunakan untuk mendapatkan hasil yang sangat akurat.

Tujuan penelitian ini bertujuan untuk mengetahui efektivitas dari implementasi algoritma Firefly-Ensemble untuk memprediksi toksisitas dengan studi kasus AR-LBD. Firefly Algorithm terbukti lebih efisien daripada algoritma optimisasi gerombolan partikel dan genetika, dan cocok untuk masalah optimisasi non-linear, unimodal, dan multimodal [15]. Kami juga menggunakan metode Ensemble untuk mengoptimalkan parameter penelitian. Metode Ensemble membuat prediksi akhir dengan pemungutan suara mayoritas dari semua hasil model individu [16]. Diharapkan penerapan algoritma Firefly-Ensemble dapat menghasilkan prediksi toksisitas pada AR-LBD, yang dapat digunakan untuk mencegah terjadinya kegagalan obat.

Topik dan Batasannya

Penelitian ini berfokus pada penerapan algoritma Firefly-Ensemble dalam memprediksi toksisitas dengan studi kasus toksisitas *Androgen Receptor Ligand Binding Domain* (AR-LBD). Seiring dengan meningkatnya kebutuhan akan metode prediksi toksisitas yang efisien dan akurat dalam pengembangan obat, pendekatan komputasi seperti *machine learning* menjadi solusi alternatif dibandingkan metode *konvensional in vitro* dan *in vivo* yang memerlukan biaya tinggi serta memiliki keterbatasan sumber daya. Dalam penelitian ini, dilakukan eksplorasi terhadap efektivitas Firefly Algorithm sebagai metode seleksi fitur, yang bertujuan untuk meningkatkan performa model dengan menghilangkan fitur yang kurang relevan dalam proses prediksi toksisitas. Selain itu, teknik Ensemble diterapkan untuk mengoptimalkan parameter model sehingga dapat meningkatkan akurasi prediksi. Topik utama yang dibahas meliputi evaluasi performa metode Firefly Algorithm dalam seleksi fitur, analisis efektivitas pengoptimalan parameter menggunakan teknik Ensemble, serta pengukuran performa keseluruhan dari kombinasi Firefly-Ensemble dalam proses prediksi toksisitas AR-LBD.

Tujuan

Penelitian ini bertujuan untuk mengimplementasikan Firefly Algorithm sebagai metode seleksi fitur dalam analisis data toksisitas AR-LBD. Proses ini dilakukan untuk mengidentifikasi fitur-fitur yang paling relevan, sehingga dapat meningkatkan efisiensi dan akurasi prediksi. Selain itu, penelitian ini juga bertujuan untuk menggunakan metode Ensemble dengan model Random Forest, AdaBoost, dan XGBoost dengan pengoptimalan parameter untuk meningkatkan kinerja model prediksi. Pengoptimalan ini melibatkan penyesuaian parameter, untuk mendapatkan hasil yang optimal berdasarkan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Penelitian ini diharapkan mampu mengevaluasi efektivitas kombinasi Firefly Algorithm dan Ensemble dalam memprediksi toksisitas pada kasus AR-LBD. Hasil yang diharapkan adalah model prediksi toksisitas yang lebih akurat dan efisien, mendukung pengembangan metode *in silico* dalam toksikologi dan penemuan obat.

Tabel 1 Keterkaitan antara tujuan, pengujian dan Kesimpulan

No.	Tujuan	Pengujian	Kesimpulan
1.	Mengimplementasikan metode Firefly Algorithm untuk melakukan seleksi fitur.	Menguji akurasi model sebelum dan sesudah seleksi fitur menggunakan Firefly Algorithm pada Random Forest, AdaBoost, dan XGBoost.	Seleksi fitur dengan Firefly Algorithm memberikan hasil yang bervariasi: 1. Random Forest: Akurasi sedikit menurun dari 0,9885 menjadi 0,9860. 2. AdaBoost: Akurasi meningkat dari 0,8263 menjadi 0,8475. 3. XGBoost: Akurasi sedikit menurun dari 0,9937 menjadi 0,9891.
2.	Menggunakan metode Ensemble untuk menetapkan parameter optimal	Menguji performa masing-masing model Ensemble dengan parameter default dan parameter optimal untuk menilai peningkatan performa.	Hyperparameter Tuning meningkatkan performa model, terutama pada AdaBoost: 1. Random Forest: Akurasi meningkat dari 0,9860 menjadi 0,9863. 2. AdaBoost: Akurasi meningkat dari 0,8475 menjadi 0,8761. 3. XGBoost: Akurasi meningkat dari 0,9814 menjadi 0,9855.
3.	Mengetahui algoritma Firefly-Ensemble bekerja dalam proses prediksi toksisitas dengan studi kasus AR-LBD	Membandingkan kinerja model Firefly-Ensemble dengan model tanpa seleksi fitur dan Hyperparameter Tuning berdasarkan metrik evaluasi (<i>accuracy</i> , <i>precision</i> , <i>recall</i> , dan <i>F1-score</i>).	Evaluasi menunjukkan bahwa kombinasi Firefly Algorithm dan Ensemble menghasilkan performa terbaik pada model XGBoost, dengan akurasi 99,30%, F1-score 99,31%, recall 99,88%, dan precision 98,75% pada data uji. Model ini diikuti oleh Random Forest dengan akurasi 99,12%, dan AdaBoost dengan akurasi 88,67%.

Organisasi Tulisan

Penyusunan tulisan dimulai dengan tinjauan literatur yang membahas berbagai algoritma seleksi fitur dan metode Ensemble yang relevan dengan penelitian ini. Selanjutnya, penjelasan tentang metodologi yang digunakan, termasuk penerapan Firefly Algorithm untuk seleksi fitur dan penggunaan model Ensemble (Random Forest, AdaBoost, dan XGBoost), akan dijabarkan. Bagian berikutnya akan membahas hasil evaluasi eksperimen, termasuk analisis perbandingan akurasi dan performa masing-masing model. Selanjutnya, jurnal ini akan menyimpulkan temuan penelitian dan memberikan rekomendasi untuk pengembangan lebih lanjut dalam penelitian tentang prediksi toksisitas menggunakan Ensemble.

2. Studi Terkait

2.1 Literatur Survey

Pada tahun 2021, Feng, et al. menggunakan metode Ensemble dan sidik jari molekuler untuk memprediksi toksisitas reproduksi zat kimia. Model Ensemble-Top12 memiliki akurasi prediksi terbaik sebesar 86,33% [10]. Penelitian yang dilakukan oleh Riwandy, et al. pada tahun 2024 menggunakan *Artificial Neural Networks* (ANNs) yang dioptimalkan oleh *Cuckoo Search Algorithm* (CSA) dan