
Pemodelan Prediktif Toksisitas Uji Klinis dengan Menggunakan Metode Cuckoo Search-Ensemble

Rafi Baihaqi¹, Isman Kurniawan²

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

⁴Divisi Digital Service PT Telekomunikasi Indonesia

¹rafibaihaqi@students.telkomuniversity.ac.id,

²ismankrn@telkomuniversity.ac.id

Abstrak

Prediksi toksisitas dalam pengembangan obat sangat penting untuk menentukan kesesuaian klinis senyawa baru. Studi saat ini meningkatkan akurasi prediksi toksisitas dengan mengintegrasikan metode ensemble, khususnya *Random Forest* dan *AdaBoost*, dengan algoritma *Cuckoo Search* untuk proses seleksi fitur yang dioptimalkan. Pendekatan baru ini diuji menggunakan dataset Clintox, yang terdiri dari 1431 senyawa kimia, yang bertujuan untuk memprediksi toksisitas klinis. Penggunaan model *Cuckoo Search-Ensemble* menunjukkan peningkatan yang signifikan dalam akurasi prediksi, yang menunjukkan ketahanan terhadap variasi data. Efektivitas model ini semakin dipertegas dengan pemilihan fitur yang ketat dan penyetelan hiperparameter, yang sangat penting untuk meningkatkan kinerja. Evaluasi model menggunakan metrik kinerja penting termasuk akurasi, presisi, *recall*, dan *F1-score*. Hasil penelitian kami menunjukkan kinerja yang baik untuk model *AdaBoost* dengan menggunakan *learning rate* 0,1 dan 150 *estimator*, di mana model ini mencapai *F1-Score* 0,6515 dan akurasi 0,9125.

Kata kunci: pencarian cuckoo, Metode *Ensemble*, Pembelajaran Mesin, Toksisitas Uji Klinis, Pemilihan Fitur, Pemodelan Prediktif

Abstract

Toxicity prediction in drug development is pivotal to determine the clinical suitability of new compounds. The current study enhances the accuracy of toxicity prediction by integrating an ensemble method, specifically *Random Forest* and *AdaBoost*, with the *Cuckoo Search* algorithm for optimized feature selection. This novel approach was tested using the Clintox dataset, comprising 1431 chemical compounds, aiming to predict clinical toxicity. The use of the *Cuckoo Search-Ensemble* model demonstrated a significant improvement in prediction accuracy, showcasing robustness against data variations. The model's effectiveness is further underscored by rigorous feature selection and hyperparameter tuning, which are crucial for enhancing performance. The evaluation of the model utilized critical performance metrics including accuracy, precision, recall, and *F1-score*. Our results found good performance for the *AdaBoost* model using a learning rate of 0.1 and 150 estimators, where it achieved an *F1-Score* of 0.6515 and an accuracy of 0.9125.

Keywords: Cuckoo Search, Ensemble Method, Machine Learning, Clinical Trial Toxicity, Feature Selection, Predictive Modeling
