

## 1. Pendahuluan

### Latar Belakang

Obat membantu menyembuhkan penyakit, tetapi seringkali menyebabkan efek samping yang tidak diinginkan atau reaksi yang merugikan. Akibat efek samping ini, banyak obat yang sebelumnya disetujui harus ditarik dari pasaran [1]. Hal ini bisa terjadi pada segala jenis obat, termasuk obat bebas atau obat yang diresepkan oleh dokter. Masalah efek samping obat menjadi perhatian serius dalam industri farmasi dan bidang kesehatan. Beberapa efek samping dapat berdampak negatif, meskipun ada juga yang bisa ditanggung oleh tubuh [2]. Efek samping obat dapat berpotensi merusak organ tubuh, dan dalam kasus tertentu, pasien bahkan dapat mengalami kematian akibat reaksi obat yang berisiko. Oleh karena itu, sangat penting untuk memastikan bahwa obat-obatan yang digunakan efektif dalam pengobatan. Karena itu, respons berbahaya terhadap obat adalah penyebab utama kegagalan obat, dan banyak obat harus ditarik dari pasar karena efek samping yang berpotensi membahayakan [3].

Menurut Centers for Disease Control and Prevention (CDC), lebih dari 2 juta individu di Amerika Serikat mengalami reaksi yang tidak diinginkan akibat obat setiap tahunnya, dan lebih dari 100.000 kematian terkait dengan situasi tersebut [4]. Contohnya, pada tahun 2008, Departemen Industri Prancis memproyeksikan bahwa hanya 1 dari 250 obat yang mendapatkan persetujuan dari FDA [2]. Banyak efek samping yang tidak terdeteksi selama proses pengembangan, penemuan, dan desain obat selama uji klinis [5]. Pengenalan dan deteksi awal terhadap efek samping obat adalah suatu permasalahan yang krusial dalam pengembangan dan keselamatan obat. Mendeteksi dan mengidentifikasi efek samping obat menjadi sangat penting untuk memastikan bahwa hanya obat-obat yang aman yang dapat diizinkan beredar dan digunakan di pasaran. Hal ini adalah upaya mendasar untuk melindungi kesehatan masyarakat dan meminimalkan risiko yang dapat timbul akibat penggunaan obat yang berpotensi berbahaya [6].

Saat ini, identifikasi efek samping obat masih memanfaatkan pengujian klinis. Karena durasi yang terbatas dan populasi pasien yang kecil, uji klinis tidak memiliki kekuatan statistik yang diperlukan untuk mengidentifikasi statistik yang parah [7]. Proses pengembangan dan desain obat melibatkan serangkaian uji klinis yang perlu dilakukan secara bertahap. Proses ini dianggap sebagai tahapan yang monoton, kompleks, sulit, dan membutuhkan investasi finansial yang besar. Dalam proses ini, diperlukan keahlian teknologi, kemampuan penelitian yang kuat, tenaga kerja terampil, senyawa kimia, dan biaya yang mencapai miliaran dolar [3]. Dalam hal ini alternatif yang dapat dilakukan adalah menggunakan machine learning untuk memprediksi efek samping obat.

Dalam beberapa tahun terakhir, teknik machine learning telah digunakan untuk memprediksi efek samping obat karena kemampuannya dalam mengelola data yang kompleks [1]. Pada tahun 2023, Xujun Liang et al., Mereka mengusulkan metode Graph Neural Network (GNN). Hasil evaluasi menunjukkan rata-rata AUC 0,8874 dan rata-rata AP 0,9200 jika hanya target dari *Dataset* DrugBank. Namun, jika hubungan obat-protein dari *Dataset* DrugBank dan STITCH dimasukkan, rata-rata AUC meningkat menjadi 0,9273 dan rata-rata AP meningkat menjadi 0,9445 [8]. Pada tahun 2015, Wen Zhang et al., menggunakan metode multi-label learning dan *ensemble* learning untuk kasus yang serupa. Metode tersebut diterapkan ketika menguji *Dataset* Liu, Kemudian, model *ensemble* tersebut menghasilkan rata-rata akurasi sebesar 0,5134 [1]. Pada tahun 2022, M.A. Arshed, et al., menggunakan metode deep learning yang dinamakan DLMSE untuk memprediksi efek samping obat berdasarkan struktur kimia obat. Penelitian sebelumnya oleh M.A. Arshed. et al., mendapatkan tingkat akurasi sebesar 0,9494 dengan menggunakan metode ini [9]. Pada tahun 2017, G.M. Dimitri, et al., Mengusulkan metode DrugClust dengan algoritma clustering untuk memprediksi efek samping obat berdasarkan fitur target protein, sub struktur kimia, dan gabungan keduanya. Hasil pengujian pada *Dataset* Zhang menunjukkan nilai AUPR (area under precision recall curve) sekitar 0,311 [10] Pada tahun 2023, R.R. Rizwandy et al., menggunakan Cuckoo Search Algorithm(CSA) untuk mengoptimalkan arsitektur Jaringan Syaraf Tiruan (JST) dalam prediksi toksisitas. Model terbaik mencapai akurasi 0,9652 dan F1 Score 0,6153, menunjukkan efektivitas CSA dalam meningkatkan performa prediksi [11].

Tantangan saat ini dalam memprediksi efek samping obat adalah kompleksitas tinggi akibat jumlah fitur yang besar. Untuk mengatasi permasalahan ini, seleksi fitur dapat digunakan sebagai alternatif. Belum banyak penelitian yang menerapkan metode seleksi fitur dalam pengembangan machine learning untuk prediksi efek samping obat. Seleksi fitur dengan metode heuristic seperti *Cuckoo Search Algorithm* bisa menjadi sebuah solusi pendekatan yang dapat digunakan.

Tujuan dari penelitian ini adalah untuk melakukan prediksi efek samping dari obat dengan

menggunakan metode *Cuckoo search-ensemble* dengan studi kasus Eye disorders. Metode cuckoo search adalah algoritma optimasi yang dikembangkan oleh Xin-She Yang dan Suash Deb (2009). Algoritma ini terinspirasi oleh perilaku burung cuckoo dalam mencari tempat untuk menetas telur [12]. Algoritma cuckoo search bekerja dengan cara mengevaluasi solusi-solusi yang dihasilkan oleh populasi telur dan memilih solusi-solusi terbaik untuk dimasukkan ke dalam populasi induk [12]. Penerapan metode cuckoo search dalam seleksi fitur untuk prediksi efek samping obat mampu mengatasi kompleksitas tinggi yang disebabkan oleh jumlah fitur yang besar [13]. Penelitian ini menggunakan *ensemble* learning pada metode cuckoo search untuk mengoptimalkan parameter terkait efek samping obat, meningkatkan akurasi klasifikasi, dan menghasilkan prediksi yang lebih akurat.

### Rumusan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana efektivitas Cuckoo Search Algorithm untuk digunakan sebagai seleksi fitur?
2. Bagaimana efektivitas optimasi hyperparameter tuning menggunakan metode Ensemble?
3. Bagaimana hasil performa model yang digunakan untuk memprediksi efek samping obat dengan metode Cuckoo search-ensemble?

### Tujuan

Tujuan penelitian ini adalah, sebagai berikut:

1. Mengimplementasikan serta mengevaluasi hasil kinerja metode *Cuckoo Search Algorithm* untuk seleksi fitur.
2. Menganalisis pengaruh dan optimasi hyperparameter terhadap performa model prediksi efek samping obat.
3. Mengevaluasi performa model prediksi efek samping obat dalam menghasilkan prediksi yang akurat menggunakan metode *Cuckoo search-ensemble*.

### Organisasi Tulisan

Organisasi tulisan pada laporan penelitian ini adalah sebagai berikut:

1. Pada sub studi terkait peneliti melakukan tinjauan pustaka terhadap beberapa paper yang menjadi referensi. Kemudian, mencari tinjauan untuk teori mengenai, perkembangan penggunaan machine learning dalam prediksi efek samping obat, penggunaan metode *Cuckoo Search Algorithm* (CSA) dalam prediksi efek samping obat, dan metode *Ensemble* dalam prediksi efek samping obat.
2. Pada sub sistem yang dibangun berisikan perancangan model prediksi yang akan dibangun. Metode heuristik yaitu *Cuckoo Search Algorithm* (CSA) akan digunakan sebagai metode untuk seleksi fitur. Kemudian model akan dibangun menggunakan metode *Ensemble* dengan tiga kernel yaitu Random Forest, AdaBoost dan XGBoost. Kemudian pada sub-bagian akan mendeskripsikan *Dataset* yang digunakan, proses seleksi fitur, pengembangan, hingga evaluasi pada model.
3. Pada sub evaluasi dijelaskan hasil pengujian dan analisis dari model prediksi yang dibangun. Pengujian dilakukan dengan hyperparameter tuning untuk meningkatkan akurasi dan menentukan parameter terbaik. Dari ketiga parameter tersebut, didapatkan nilai best parameter yang lebih baik dari nilai default parameter. Evaluasi masing masing hasil set pelatihan dan set pengujian dari ketiga kernel dilakukan menggunakan *confusion matrix*. Setelah itu, dilakukan analisis terhadap hasil performa set pelatihan dan set pengujian dari ketiga kernel berdasarkan beberapa *matrix*, yaitu Accuracy, Recall, Precision, dan F1-Score.
4. Pada sub kesimpulan berisikan kesimpulan dari penelitian yang telah dilakukan, dan saran untuk penelitian yang akan dilakukan pada masa mendatang