

1. Pendahuluan

1.1. Latar Belakang

Manusia sangat berpotensi untuk terpapar berbagai jenis senyawa kimia yang diantaranya berpotensi berbahaya [1], [2]. Oleh karena itu penentuan toksisitas suatu bahan kimia sangat penting untuk meminimalisir paparan terhadap zat berbahaya dalam produk sehari-hari [1]. Selain itu toksisitas merupakan isu sentral dalam pengembangan obat baru, dimana lebih dari 30% kandidat obat gagal dalam uji klinis karena tidak diketahui efek racun untuk tubuhnya [1], [3]. Pada tahun 2008, the U. S. National Institutes of Health (NIH) and the U. S. Environmental Protection Agency (EPA), telah berkolaborasi untuk pengujian toksisitas di masa depan [4]. Upaya mereka diikuti oleh U.S. Food and Drug Administration (FDA) di bawah naungan Tox21 Program [2]. Program tersebut bertujuan untuk mengembangkan metode penilaian toksisitas yang lebih baik, hal ini dilakukan karena metode saat ini kurang memadai seiring dengan meningkatnya permintaan untuk pengujian toksisitas yang efektif [4].

Metode yang digunakan untuk menguji toksisitas dalam jumlah besar adalah *High-Throughput Screening* (HTS). Eksperimen ini meneliti apakah suatu senyawa kimia pada kadar tertentu menunjukkan adanya jenis toksisitas tertentu [5]. Percobaan ini dilakukan secara berulang dengan menggunakan berbagai konsentrasi senyawa kimia yang memungkinkan untuk menentukan kurva dosis konsentrasinya. Kekurangan dari metode HTS adalah proses yang lama dan biaya yang besar [2]. Selain itu juga metode ini memiliki kekurangan dalam hal diagnosis terkait tidak terdeteksinya varian yang berupa non-coding (berlaku jika tidak ditarget) dan analisis yang rumit [6]. Dikarenakan kekurangan tersebut maka dibutuhkan metode yang lebih akurat dan tidak perlu memakan waktu dan biaya yang banyak. Salah satu alternatif metode dapat digunakan adalah implementasi *machine learning* untuk mendapatkan hasil yang lebih akurat dan tidak memakan waktu yang lama dan biaya yang besar. Di antara semua fitur biometrik yang dapat digunakan untuk identifikasi *fingerprints* merupakan salah satu jenis fitur yang paling banyak digunakan [7], [8].

Penelitian mengenai Implementasi *Machine Learning* untuk prediksi toksisitas berbasis *fingerprint* sudah beberapa kali dilakukan sebelumnya. Pada tahun 2018 Fan dan koleganya melakukan penelitian prediksi genotoksitas kimia menggunakan *machine learning* dan peringatan sturatural. Pada penelitian ini didapatkan hasil akurasi menggunakan MACCS *fingerprints* dan menggunakan metode *Random Forest* didapatkan hasil 94.7% [9]. Pada tahun 2020 Feng dan koleganya melakukan penelitian untuk memprediksi tingkat reproduksi toksisitas dengan menggunakan data molekuler *fingerprint* dan didapat hasil akurasinya 86.33% dengan menggunakan *Support Vector Machine*, *Random Forest*, *Extream Gradient Boosting Method* [10]. Pada tahun 2021 Xu dan koleganya melakukan penelitian membandingkan enam metode *machine learning* yang dikombinasikan dengan sembilan *fingerprint* molekul untuk menetapkan 54 model klasifikasi biner untuk prediksi toksisitas kontak akut pada lebah madu [11]. Dari hasil penelitian tersebut didapatkan nilai paling akurat adalah menggunakan metode *Support Vector Machine* dengan dikombinasikan dengan *fingerprint* diperpanjang CDK dengan nilai akurasi 0,04% [11].

Pada tahun 2021 Su dan koleganya melakukan penelitian mencari metode penyeimbangan data yang efektif untuk genotoksitas dataset imbalanced menggunakan *machine learning* dan *fingerprint* molekul. Dari hasil penelitian tersebut didapatkan nilai akurasi paling tinggi terdapat pada penggunaan dataset molekul MACCS dan dengan penyeimbang data menggunakan *Synthetic Minority Oversampling Technique* (SMOTE) dan menggunakan algoritma *Gradient Boosting Tree* (GBT) dengan hasil 66,7% [12].

Berdasarkan literatur survey yang dilakukan untuk mengidentifikasi kadar toksisitas pada data *fingerprint* menggunakan *machine learning*. Penentuan arsitektur *Artificial Neural Network* (ANN) masih banyak dilakukan dengan manual tuning. Untuk alternatif dari penggunaan manual tuning untuk mendapatkan hasil yang lebih akurat dapat digunakan auto tuning menggunakan *meta-heuristic*. Salah satu metode *meta-heuristic* yang dapat digunakan adalah *Cuckoo Search Algorithm*.

Penelitian ini bertujuan untuk mengimplementasi *Cuckoo Search Algorithm* pada Optimasi Metode *Artificial Neural Network* untuk Prediksi Toksisitas berbasis Fitur *Fingerprint*. *Cuckoo Search Algorithm* adalah *meta-heuristic search algorithm* yang terinspirasi dari reproduksinya burung kukuk. Dimana setiap burung kukuk ingin bertelur akan dipilih secara acak burung lain dan sarang dari kualitas telur terbaik akan digunakan untuk reproduksi selanjutnya [13].

1.2. Rumusan Masalah

Rumusan masalah pada penelitian ini adalah:

1. Bagaimana efektifitas model baseline menggunakan metode *Artificial Neural Network* ?
2. Bagaimana pengaruh model arsitektur menggunakan metode *Cuckoo Search Algorithm* ?

3. Bagaimana hasil performa prediksi menggunakan metode *Cuckoo Search Algorithm* dan *Artificial Neural Network (ANN)* ?

1.3. Tujuan

Tujuan pada penelitian ini adalah :

1. Mengetahui efektifitas model baseline pada metode *Artificial Neural Network*.
2. Mengetahui pengaruh model arsitektur menggunakan metode *Cuckoo Search Algorithm*.
3. Mengetahui hasil performa prediksi menggunakan metode *Cuckoo Search Algorithm* dan *Artificial Neural Network*.