

# Implementasi Firefly Algorithm-Support Vector Machine dalam Mengklasifikasikan Toksisitas Uji Klinis

Raden Ananta Mahardika<sup>1</sup>, Isman Kurniawan<sup>2</sup>

<sup>1,2</sup>Fakultas Informatika, Universitas Telkom, Bandung

<sup>1</sup>radram@student.telkomuniversity.ac.id, <sup>2</sup>ismankrn@telkomuniveristy.ac.id

---

## Abstrak

Obat-obatan memainkan peran penting dalam meningkatkan kesehatan dan kesejahteraan manusia. Sebelum disetujui untuk digunakan oleh publik, obat-obatan harus menjalani uji klinis untuk menilai toksisitasnya dan memastikan keamanannya. Namun, metode konvensional, seperti pengujian pada hewan, mahal, memakan waktu, kontroversial secara etis, dan membutuhkan banyak tenaga. Untuk mengatasi tantangan ini, pendekatan alternatif untuk prediksi toksisitas sangat penting. Penelitian ini berfokus pada Support Vector Machine (SVM), sebuah algoritma Machine Learning (ML) yang diawasi, yang sangat efektif untuk tugas-tugas klasifikasi. Kernel yang digunakan dalam penelitian ini adalah linear, poly, dan RBF. Selain itu, Firefly Algorithm (FA) digunakan sebagai teknik optimasi untuk pemilihan fitur, yang bertujuan untuk meningkatkan akurasi model dan mengurangi dimensi data. Penelitian ini mengeksplorasi implementasi pendekatan hibrida FA-SVM untuk memprediksi toksisitas obat menggunakan dataset yang berisi representasi SMILES, persetujuan FDA, dan label toksisitas. Penelitian ini membandingkan akurasi yang dicapai oleh kernel SVM yang berbeda, dengan hasil yang menunjukkan bahwa kernel linear mencapai F1-Scores 0,987, kernel poly 0,969, dan kernel RBF 0,976.

**Kata kunci:** machine learning, support vector machine, obat, firefly algorithm.

---

## Abstract

Drugs play a crucial role in improving human health and well-being. Before being approved for public use, drugs must undergo clinical trials to assess their toxicity and ensure safety. However, conventional methods, such as animal testing, are expensive, timeconsuming, ethically controversial, and laborintensive. To address these challenges, alternative approaches for toxicity prediction are essential. This study focuses on the Support Vector Machine (SVM), a supervised Machine Learning (ML) algorithm, which is highly effective for classification tasks. The kernels used in this study are linear, poly, and RBF. Additionally, the Firefly Algorithm (FA) is employed as an optimization technique for feature selection, aiming to improve model accuracy and reduce data dimensionality. The study explores the implementation of the FA-SVM hybrid approach to predict drug toxicity using a dataset containing SMILES representations, FDA approval and toxicity labels. The study compares the accuracy achieved by the different SVM kernels, with results indicating that the linear kernel achieved an F1-Scores of 0.987, the poly kernel 0.969, and the RBF kernel 0.976. **Keywords:** keyword should be chosen that they best describe the contents of the paper and should be typed in lower-case, except abbreviation.

**Keywords:** machine learning, support vector machine, drugs, firefly algorithm.

---

## 1. Pendahuluan

### 1.1. Latar Belakang

Obat adalah bahan atau paduan bahan yang digunakan untuk mempengaruhi atau menyelidiki sistem fisiologi atau keadaan patologi dalam rangka penetapan diagnosis, pencegahan, penyembuhan, pemulihan, peningkatan kesehatan dan kontrasepsi untuk manusia. Namun, walaupun obat memiliki potensi besar dalam penyembuhan dan pencegahan penyakit, obat juga dapat menimbulkan efek samping dan toksisitas pada manusia. Oleh karena itu, uji coba klinis dan toksisitas dilakukan pada obat sebelum dijual ke masyarakat. Toksisitas atau tingkat adalah tingkat merusaknya suatu zat jika dipaparkan terhadap organisme. Uji coba dan evaluasi toksisitas merupakan hal yang sangat penting dalam pengembangan obat, karena membantu meningkatkan keamanan dalam mengkonsumsi obat. Pada tahun 2022, proses pengembangan obat terdapat kemungkinan gagal sebesar 90% [1]. Dari 90% kemungkinan kegagalan pengembangan obat tersebut ada empat kemungkinan faktor yang terkait yaitu: kurangnya kemanjuran klinis (40%-50%), tidak terkendalinya toksisitas (30%), properti *drug-like* yang buruk (10%-15%) dan perencanaan strategi yang buruk (10%) [2]. Dari data tersebut dapat disimpulkan bahwa toksisitas menjadi salah satu faktor yang membuat pengembangan obat menjadi gagal dan tidak dapat diedarkan kepada masyarakat.

Dalam pembuatan dan pengembangan obat terdapat cara konvensional dalam melakukan pengujian obat salah satu metodenya yaitu melakukan uji coba obat kepada hewan. Akan tetapi, terdapat kekurangan dalam melakukan uji

coba obat terhadap hewan yaitu, menggunakan biaya yang sangat mahal dan sangat menguras tenaga dalam proses pengerjaannya dan hasil uji coba tersebut tidak dapat memberikan informasi yang cukup untuk mengindikasikan bahwa obat tersebut aman untuk dikonsumsi manusia [3]. Untuk mengurangi biaya dan resiko yang ada pada metode uji coba terhadap hewan, sangat penting untuk melakukan uji coba prediksi dengan cara alternatif. Salah satu metode prediksi toksisitas yang dominan dan paling dikembangkan adalah metode *in silico Quantitative Structure-Activity Relationship (QSAR)* berdasarkan parameter struktur bahan kimia [4]. Pada tahun 2020, Rim membahas tentang prediksi toksisitas *in silico* dan pengaplikasiannya. Dalam bahasannya dia menjelaskan bahwa metode *in silico* dapat memprediksi toksisitas zat kimia, campuran, dan *nanomaterial* secara kuantitatif dan kualitatif. Teknik ini juga dapat digunakan untuk mempercepat proses pengambilan keputusan dalam penilaian risiko kimia [5].

Terdapat beberapa studi penelitian yang berkaitan prediksi klinis toksisitas yang menggunakan *Machine Learning (ML)* yang sudah pernah dilakukan sebelumnya. Pada tahun 2021, Jaganathan dan rekan melakukan klasifikasi biner dengan model *Support Vector Machine (SVM)*, *Random Forest (RF)* dan beberapa model lainnya untuk memprediksi toksisitas hati yang terpengaruh karena obat-obatan. Pada model SVM mendapatkan skor akurasi sebesar 0.811 atau sebesar 81,1% dan pada model *Random Forest* mendapatkan skor akurasi sebesar 0.768 atau 76,8% [6]. Pada tahun 2014, Zhou dan rekan membangun model ML untuk memprediksi toksisitas ototoksik yang diinduksi oleh obat. Mereka menggunakan model SVM dan mendapatkan nilai akurasi sebesar 85.33% [7]. Pada tahun 2015, Cao dan rekan melakukan prediksi toksisitas bahan kimia berdasarkan *database EPA toxicity* menggunakan pendekatan *kernel fusion* dengan SVM. Dalam hasil prediksi, mereka mendapatkan tingkat akurasi sampai 93.6% [8]. Pada tahun 2015, Chao dan Horng melakukan percobaan klasifikasi menggunakan SVM yang menggunakan FA dan SVM tanpa FA. Dari hasil percobaan tersebut mereka menggunakan fitur optimasi yang disediakan dari FA sehingga keakuratan model SVM yang menggunakan FA mendapatkan akurasi lebih besar dibandingkan SVM tanpa menggunakan FA yaitu sebesar 92.5% [9]. Pada tahun 2022, Cavasotto dan Scardino membahas beberapa model ML untuk memprediksi toksisitas salah satunya adalah model SVM. Pada bahasannya SVM digunakan dalam kombinasi dengan *MACCS keys* dan *fingerprinting* untuk memperoleh hasil yang sangat baik dalam memprediksi toksisitas hepatik. Model SVM yang mereka buat untuk memprediksi mendapatkan tingkat akurasi sebesar 80.4% [10]. Pada tahun 2018, Wu dan Wang menggunakan model SVM untuk melakukan klasifikasi bahan kimia berdasarkan toksisitasnya dan mendapatkan nilai AUC rata-rata sebesar 91% [3].

Dari beberapa penelitian yang sudah dilakukan dapat disimpulkan bahwa metode memprediksi toksisitas dengan model SVM sudah mendapatkan hasil akurasi yang baik. Akan tetapi, masih dapat ditingkatkan akurasinya dan model SVM memberikan hasil yang memungkinkan untuk diimplementasikan dan mempunyai kemungkinan bisa menjadi solusi alternatif untuk memprediksi toksisitas. Dari keberagamannya akurasi model SVM di atas, model SVM dapat digabungkan dengan algoritma untuk meningkatkan kualitas fitur dari model SVM dalam memprediksi, salah satu algoritmanya adalah *Firefly Algorithm (FA)*.

FA adalah sebuah algoritma metaheuristik yang terinspirasi oleh perilaku kunang-kunang. Algoritma ini ditemukan pertama kali oleh Yang. FA memiliki keunggulan dalam mengoptimalkan parameter secara otomatis, menggunakan konsep atraksi cahaya untuk mencari solusi terbaik. Algoritma ini efisien dalam mengatasi berbagai masalah optimasi, salah satunya adalah pengelompokan data [11][12].

Penelitian ini bertujuan untuk menunjukkan bahwa uji klinis dengan metode ML dengan menggunakan model SVM dapat menjadi solusi alternatif dari metode uji klinis menggunakan hewan dan bertujuan untuk mengembangkan model SVM dalam memprediksi toksisitas dalam uji klinis dengan FA. Algoritma ini dipilih karena dapat digunakan untuk meringankan komputasi dari model SVM dengan fitur optimasi yang disediakan oleh FA. Dalam model SVM, FA dapat digunakan untuk menemukan parameter SVM secara optimal. Dari teknik tersebut penelitian ini diharapkan dapat mempercepat waktu komputasi dan meningkatkan performa model SVM dalam memprediksi toksisitas dalam uji klinis dengan menggunakan FA.

## 1.2. Topik dan Batasannya

Pembuatan dan pengembangan obat memerlukan uji coba klinis agar aman dikonsumsi oleh masyarakat. Namun, metode uji coba klinis secara konvensional memakan waktu yang cukup lama, biaya yang sangat tinggi, serta tenaga yang besar. Oleh karena itu, diperlukan metode alternatif untuk memprediksi toksisitas uji klinis dengan lebih efisien. Salah satu metode yang dapat digunakan adalah ML. Meskipun berbagai penelitian telah memanfaatkan metode ini sebagai solusi alternatif, tingkat akurasi dan performa yang dihasilkan sangat beragam tergantung pada model dan algoritma yang digunakan. Dengan demikian, penelitian lebih lanjut diperlukan untuk mengembangkan model yang mampu meningkatkan keakuratan dan efisiensi performa dalam memprediksi toksisitas uji klinis.

Perumusan masalah dalam penelitian ini berfokus pada efektivitas FA dalam meningkatkan kinerja model SVM

melalui seleksi fitur yang optimal. Selain itu, penelitian ini mempertanyakan sejauh mana *hyperparameter tuning* pada model SVM yang dioptimalkan dengan FA dapat meningkatkan akurasi dalam klasifikasi toksisitas uji klinis. Selanjutnya, penelitian ini juga akan mengevaluasi performa keseluruhan model SVM-FA dalam mengklasifikasikan toksisitas uji klinis berdasarkan metrik evaluasi yang relevan.

Penelitian ini memiliki batasan yang bertujuan untuk memfokuskan ruang lingkup studi, yaitu: pertama, penelitian ini hanya berfokus pada kombinasi model SVM dengan FA. Kedua, *dataset* yang digunakan terbatas pada *dataset* clintox.csv. Ketiga, eksperimen dilakukan hanya dengan tiga jenis *kernel* pada model SVM, yaitu *linear*, *poly*, dan RBF.

### **1.3. Tujuan**

Penelitian ini bertujuan untuk menunjukkan bahwa uji klinis menggunakan metode ML dengan model SVM dapat menjadi alternatif dari metode uji klinis yang menggunakan hewan. Selain itu, penelitian ini bertujuan untuk mengembangkan model SVM dalam memprediksi toksisitas pada uji klinis dengan memanfaatkan algoritma FA. Algoritma FA dipilih karena kemampuannya untuk meningkatkan keakuratan model SVM melalui seleksi fitur otomatis. Dalam model SVM, FA dapat digunakan untuk menemukan parameter optimal yang diperlukan untuk meningkatkan performa prediksi. Dengan teknik tersebut, penelitian ini diharapkan dapat meningkatkan keakuratan dan performa model SVM dalam memprediksi toksisitas pada uji klinis. Secara spesifik, penelitian ini bertujuan untuk: menunjukkan efektivitas FA sebagai metode optimasi seleksi fitur pada model SVM, menguji efektivitas *hyperparameter tuning* pada kombinasi SVM-FA dalam memprediksi toksisitas uji klinis, dan mengevaluasi performa model SVM-FA dalam memprediksi toksisitas pada uji klinis.

### **1.4. Organisasi Tulisan**

Setelah memaparkan pendahuluan pada bagian pertama, bagian kedua akan membahas kajian terkait SVM FA dan SMILES, termasuk perhitungan performansinya. Pada bagian ketiga, fokus akan beralih ke sistem yang dikembangkan, termasuk deskripsi *dataset* dan analisis performansi. Bagian keempat akan menyajikan evaluasi, meliputi skenario pengujian, hasil pengujian, dan analisis dari hasil pengujian tersebut. Terakhir, bagian kelima akan menyajikan kesimpulan dan saran dari penelitian ini.