

PERBANDINGAN MODEL *MACHINE LEARNING* UNTUK *VIRTUAL SCREENING* SENYAWA KANDIDAT PENGGANTI AGP (*ANTIBIOTIC GROWTH PROMOTER*)

AINUN FADHILA AZ ZAHIDA



**PROGRAM STUDI ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Perbandingan Model *Machine Learning* untuk *Virtual Screening* Senyawa Kandidat Pengganti AGP (*Antibiotic Growth Promoter*)” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Desember 2025

Ainun Fadhila Az Zahida
G64190002



ABSTRAK

AINUN FADHILA AZ ZAHIDA. Perbandingan Model *Machine Learning* untuk *Virtual Screening* Senyawa Kandidat Pengganti AGP (*Antibiotic Growth Promoter*). Dibimbing oleh LAILAN SAHRINA HASIBUAN.

Penggunaan *Antibiotic Growth Promoter* (AGP) dalam industri unggas telah dibatasi karena kontribusinya terhadap meningkatnya resistensi antibiotik, sehingga diperlukan upaya pencarian kandidat antibiotik alternatif yang aman dan efektif. Penelitian ini bertujuan membangun dan membandingkan performa model klasifikasi *Support Vector Machine* (SVM), C5.0, dan *Extreme Gradient Boosting* (XGBoost) dalam merekomendasikan kandidat senyawa antibiotik alternatif AGP, serta mengidentifikasi fitur potensial yang berperan dalam prediksi aktivitas senyawa. Pendekatan *virtual screening* dilakukan secara *in silico* menggunakan dataset yang terdiri atas 2.335 senyawa, dengan 120 senyawa aktif dan 2.215 senyawa inaktif. Sebanyak 261 deskriptor molekul diekstraksi dari representasi SMILES (*Simplified Molecular Input Line Entry System*) menggunakan paket rcdk, diikuti penanganan ketidakseimbangan data dengan ADASYN dan seleksi fitur berbasis *Near Zero Variance*. Model dilatih menggunakan skema *5-fold cross-validation* dan dievaluasi menggunakan metrik akurasi, *recall*, *precision*, *F1-Score*, dan PR-AUC. Hasil menunjukkan bahwa ketiga model mencapai akurasi tinggi pada data uji (93,62-96,58%), namun sensitivitas terhadap kelas minoritas masih terbatas dengan nilai *recall* 0,30-0,35. Di antara ketiga model, XGBoost menunjukkan performa paling stabil dengan nilai PR-AUC tertinggi sebesar 0,52. Analisis fitur menunjukkan bahwa deskriptor *topological* dan *hybrid* berperan dominan dalam proses klasifikasi yang menegaskan pentingnya representasi struktur molekul dalam prediksi aktivitas antibakteri.

Kata kunci: AGP, *machine learning*, rcdk, SMILES, *virtual screening*

ABSTRACT

AINUN FADHILA AZ ZAHIDA. Comparative Study of Machine Learning Models for Virtual Screening of Alternative AGP (Antibiotic Growth Promoter) Compounds. Supervised by LAILAN SAHRINA HASIBUAN.

The use of Antibiotic Growth Promoters (AGPs) in poultry production has been restricted due to their association with antimicrobial resistance, highlighting the need to identify safe and effective alternative compounds. This study aims to develop and compare the performance of Support Vector Machine (SVM), C5.0, and Extreme Gradient Boosting (XGBoost) classification models for recommending candidate compounds as AGP alternatives, as well as to identify potential features associated with compound activity. An *in silico* virtual screening approach was applied to a dataset comprising 2.335 compounds, including 120 active and 2.215 inactive compounds. A total of 261 molecular descriptors were extracted from SMILES representations using the rcdk package, followed by class imbalance handling with the ADASYN method and feature selection based on Near Zero Variance. Models were trained using a 5-fold cross-validation scheme and evaluated using accuracy, recall, precision, F1-Score, and PR-AUC metrics. The results show that all models achieved high test accuracy (93,62-96,58%), but exhibited limited sensitivity to the minority class (recall 0,30-0,35). Among the evaluated models, XGBoost demonstrated the most stable performance with the highest PR-AUC value of 0,52. Feature analysis indicates that topological and hybrid descriptors play an important role in predicting antibacterial activity.

Keywords: AGP, machine learning, rcdk, SMILES, virtual screening

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2025
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

PERBANDINGAN MODEL *MACHINE LEARNING* UNTUK *VIRTUAL SCREENING* SENYAWA KANDIDAT PENGGANTI AGP (*ANTIBIOTIC GROWTH PROMOTER*)

AINUN FADHILA AZ ZAHIDA

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Ilmu Komputer

**PROGRAM STUDI ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Skripsi:

- 1 Hari Agung Adrianto, S.Kom., M.Si., Ph.D.
- 2 Muhammad Asyhar Agmalaro, S.Si., M.Kom.

Judul Skripsi : Perbandingan Model *Machine Learning* untuk *Virtual Screening*
Senyawa Kandidat Pengganti AGP (*Antibiotic Growth Promoter*)

Nama : Ainun Fadhila Az Zahida

NIM : G64190002

Disetujui oleh

Pembimbing 1:

Lailan Sahrina Hasibuan, S.Kom., M.Kom.

Diketahui oleh

Ketua Program Studi Ilmu Komputer:

Dr. Sony Hartono Wijaya, S.Kom., M.Kom.

NIP 19810809 200812 1 002

Tanggal Ujian:
(27 Agustus 2025)

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah *subhanaahu wa ta'ala* atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Judul yang dipilih dalam penelitian yang dilaksanakan sejak bulan November 2022 sampai bulan Desember 2025 ini ialah “Perbandingan Model *Machine Learning* untuk *Virtual Screening* Senyawa Kandidat Pengganti AGP (*Antibiotic Growth Promoter*)”. Skripsi ini dapat tersusun atas bimbingan dan kerja sama dari berbagai pihak selama penelitian dan penyusunan skripsi. Pada kesempatan ini, penulis mengucapkan terima kasih kepada:

1. Ibu Lailan Sahrina Hasibuan., S.Kom., M.Kom. selaku dosen pembimbing akademik yang telah membimbing, memberikan arahan, dan saran kepada penulis.
2. Bapak Hari Agung Adrianto, S.Kom., M.Si., Ph.D. dan Bapak Muhammad Asyhar Agmalero, S.Si. M.Kom. selaku dosen penguji dan moderator pada sidang ujian sarjana penulis.
3. Orang tua penulis, yaitu Bapak Rochmad Tejo Kuncoro, S.TP. dan Ibu Dewi Widyastuti Yulian, S.P., S.Pd. yang selalu memberikan doa, dukungan, dan kasih sayang sehingga penelitian ini dapat terselesaikan.
4. Keluarga, adik-adik penulis, Aisy, Akhyar, dan Azma yang selalu memberikan dukungan dan semangat selama perjalanan panjang penelitian ini.
5. Shofura Qurotul ‘Aini yang selalu kebersamai penulis selama proses penelitian dan memberi dukungan secara emosional dan material.
6. Teman-teman penulis, Aisna, Iyak, Lili, Abiyya, Tri, Fathin, Alfi, dan Rere yang telah memberi doa, dukungan serta menemani dan membantu selama proses penelitian berlangsung.
7. Zhong Chenle yang telah menemani dan memberikan banyak inspirasi serta energi positif sehingga penulis dapat bertahan untuk menyelesaikan penelitian ini.
8. Teman-teman humayzen, Ambar, Thifa, Yasmina, Ayu, Tyas, Uly, dan Asma yang telah memberikan semangat dan bantuan selama penelitian.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Desember 2025

Ainun Fadhila Az Zahida

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	x
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Manfaat	2
1.5 Ruang Lingkup	2
II TINJAUAN PUSTAKA	3
2.1 AGP (<i>Antibiotic Growth Promoter</i>)	3
2.2 <i>Eschericia coli</i>	3
2.3 SMILES (<i>Simplified Molecular Input Line Entry System</i>)	3
2.4 SVM (<i>Support Vector Machine</i>)	4
2.5 C5.0	5
2.6 XGBoost (<i>Extreme Gradient Boosting</i>)	6
III METODE	8
3.1 Data Penelitian	8
3.2 Tahapan Penelitian	8
3.3 Lingkungan Pengembangan	17
IV HASIL DAN PEMBAHASAN	18
4.1 Ekstraksi Fitur	18
4.2 Praproses dan Pembagian Data	19
4.3 Penanganan Ketidakseimbangan Data	20
4.4 Seleksi Fitur	20
4.5 Pelatihan Model dan <i>Hyperparameter Tuning</i>	21
4.6 Evaluasi Model	22
4.7 Identifikasi Fitur Potensial	24
V SIMPULAN DAN SARAN	26
5.1 Simpulan	26
5.2 Saran	26
DAFTAR PUSTAKA	27
LAMPIRAN	31
RIWAYAT HIDUP	35



DAFTAR TABEL

1	Aturan dasar spesifikasi notasi SMILES (Weininger 1988)	3
2	Dataset penelitian beserta atribut	8
3	Nilai konfigurasi <i>hyperparameter</i> SVM	14
4	Nilai konfigurasi <i>hyperparameter</i> C5.0	15
5	Nilai konfigurasi <i>hyperparameter</i> XGBoost	15
6	Contoh hasil ekstraksi fitur	19
7	Hasil perhitungan seleksi fitur NZV	21
8	Metrik hasil evaluasi <i>training cross-validation</i> dan data uji	22

DAFTAR GAMBAR

1	Contoh notasi SMILES senyawa morfin (Weininger 1988)	4
2	Ilustrasi <i>hyperplane</i> (Du <i>et al.</i> 2024)	5
3	Diagram alir penelitian	9
4	Diagram alir tahapan ekstraksi fitur	11
5	Visualisasi 2D struktur graf molekul praxadine hidroklorida	18
6	Perbandingan proporsi kelas aktif dan inaktif sebelum (a) dan sesudah (b) penerapan ADASYN	20
7	Hasil <i>PR-Curve</i> ketiga model	23
8	Hasil identifikasi 10 fitur potensial pada XGboost (a) dan C5.0 (b)	25

DAFTAR LAMPIRAN

1	Implementasi algoritma ekstraksi fitur menggunakan rcdk	32
2	Contoh <i>output</i> hasil <i>parsing</i> molekul dalam objek IAtomContainer	33
3	Hasil perhitungan <i>overall</i> dan deskripsi fitur potensial pada XGBoost dan C5.0	34