



PENGEMBANGAN MODEL *DRUG - TARGET INTERACTION* UNTUK IDENTIFIKASI POTENSI SENYAWA BIOAKTIF PADA KEMUNING SEBAGAI ANTI-OBESITAS DENGAN *MACHINE LEARNING*

@Hak cipta milik IPB University

ZALANA ENDESNAH



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



PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Pengembangan Model *Drug-Target Interaction* untuk Identifikasi Potensi Senyawa Bioaktif pada Kemuning sebagai Anti-Obesitas dengan *Machine Learning* 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 tesis ini.

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

Bogor, Maret 2025

Zalana Endesnah
G6501222044

RINGKASAN

ZALANA ENDESNAH. Pengembangan Model *Drug-Target Intercation* untuk Identifikasi Potensi Senyawa Bioaktif pada Kemuning sebagai Anti-obesitas dengan *Machine Learning*. Dibimbing oleh WISNU ANANTA KUSUMA dan MUSHTHOFA.

Visceral obesity merupakan kondisi penumpukan lemak berlebih yang dapat memicu komplikasi metabolik seperti diabetes melitus tipe 2, penyakit kardiovaskular, serta peradangan kronis yang berdampak pada penurunan kualitas hidup. Salah satu pendekatan dalam mengelola obesitas adalah melalui pemanfaatan senyawa bioaktif dari tanaman herbal seperti kemuning. Kemuning (*murraya paniculata*) merupakan tanaman herbal yang secara tradisional digunakan sebagai jamu pelangsing. Ekstrak kemuning diketahui mengandung sifat metabolit sekunder seperti flavonoid, kumarin, dan terpenoid yang berperan penting dalam potensi terapeutik. Sifat metabolit sekunder pada kemuning memiliki berbagai aktivitas farmakologis, termasuk anti-inflamasi yang sering dikaitkan dengan pengelolaan obesitas.

Prediksi *drug-target interactions* (DTI) menjadi aspek penting dalam pengembangan obat modern untuk mengidentifikasi kandidat senyawa yang berpotensi menarget protein terkait obesitas dengan lebih efisien. Salah satu tantangan utama dalam penelitian DTI adalah ketidakseimbangan data, di mana jumlah interaksi terkonfirmasi jauh lebih sedikit dibandingkan interaksi yang belum diketahui. Mengatasi permasalahan tersebut, penelitian ini mengusulkan pendekatan *semi-supervised learning* menggunakan algoritma *Neighborhood Regularized Logistic Matrix Factorization* (NRLMF). Perhitungan *Cosine Similarity* diterapkan untuk menilai kemiripan antara protein dan senyawa, yang berfungsi sebagai informasi tambahan dalam pembangkitan interaksi negatif. Algoritma NRLMF menerapkan regularisasi berbasis tetangga untuk meningkatkan prediksi interaksi senyawa-protein dan membangkitkan interaksi negatif guna mengidentifikasi pasangan yang tidak berinteraksi. Data protein BAT diperoleh dari basis data OMIM dan GeneCards, yang kemudian dicari interaksi senyawa dari basis dari BindingDB. Penelitian ini membandingkan kinerja model *Random Forest* (RF), *XGBoost*, *Multi-Layer Perceptron* (MLP), dan *Logistic Regression* (LR) menggunakan interaksi negatif yang dihasilkan dari NRLMF.

Hasil penelitian menunjukkan kombinasi fitur *AAIndex_Morgan* pada model MLP memberikan performa paling optimal dari nilai akurasi, *recall*, *F1 Score*, dan Cohen Kappa, kecuali nilai *precision* dan skor AUC yang diungguli oleh model RF dengan nilai (AUC = 99,99%). Perhitungan nilai signifikansi prediksi pada MLP dan XGBoost menghasilkan tingkat signifikansi yang rendah dengan nilai *p-value* di atas 0,05. Sedangkan, RF menghasilkan nilai Z-Score dengan tingkat signifikansi paling tinggi sebesar 99%. Temuan ini menunjukkan bahwa implementasi algoritma NRLMF untuk menghasilkan interaksi negatif guna melengkapi data pemodelan *machine learning* prediksi DTI dapat menjadi metode yang lebih efektif untuk mengatasi ketidakseimbangan data.

Kata kunci: *drug-target interaction* (DTI), *imbalanced data*, kemuning, *random forest*, *semi-supervised learning*.



SUMMARY

ZALANA ENDESNAH. Development of a Drug-Target Interaction Model to Identify Potential Bioactive Compounds in Orange Jasmine as Anti-Obesity Using Machine Learning. Dibimbing oleh WISNU ANANTA KUSUMA dan MUSHTHOFA.

Visceral obesity is a condition characterized by excessive fat accumulation, which can trigger metabolic complications such as type 2 diabetes mellitus, cardiovascular diseases, and chronic inflammation, ultimately leading to a decline in quality of life. One approach to managing obesity is through the utilization of bioactive compounds from herbal plants, such as *Murraya paniculata* (kemuning). Traditionally used as a slimming herbal remedy, *Murraya paniculata* is known to contain secondary metabolites such as flavonoids, coumarins, and terpenoids, which play a crucial role in its therapeutic potential. These secondary metabolites exhibit various pharmacological activities, including anti-inflammatory properties, which are often associated with obesity management.

Drug-target interaction (DTI) prediction is a crucial aspect of modern drug development, enabling the identification of potential candidate compounds that effectively target obesity-related proteins. One of the primary challenges in DTI research is data imbalance, where the number of confirmed interactions is significantly lower than the number of unknown interactions. To address this issue, this study proposes a semi-supervised learning approach using the Neighborhood Regularized Logistic Matrix Factorization (NRLMF) algorithm. Cosine similarity calculations are applied to assess the similarity between proteins and compounds, providing additional information for generating negative interactions. The NRLMF algorithm applies neighbor-based regularization to improve compound-protein interaction prediction and generate negative interactions to identify non-interacting pairs. BAT protein data is obtained from the OMIM and GeneCards databases, while compound interaction data is sourced from the BindingDB database. This study evaluates the performance of Random Forest (RF), XGBoost, Multi-Layer Perceptron (MLP), and Logistic Regression (LR) models using negative interactions generated by NRLMF.

The results showed that the combination of AAIndex_Morgan features in the MLP model provided the most optimal performance in terms of accuracy, *recall*, F1 Score, and Cohen Kappa values, except for the *precision* and AUC scores which were outperformed by the RF model with a value (AUC = 99,99%). The calculation of the prediction significance value in MLP and XGBoost produced a low level of significance with a p-value above 0,05. Meanwhile, RF produced a Z-Score value with the highest level of significance of 99%. These findings suggest that implementing the NRLMF algorithm to generate negative interactions to complement DTI predictive machine learning modeling data can be a more effective method to address data imbalance.

Keywords: drug-target interaction (DTI), imbalanced data, kemuning, random forest, semi-supervised learning.



@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.

© Hak Cipta milik IPB, tahun 2025¹
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.



PENGEMBANGAN MODEL *DRUG - TARGET INTERACTION* UNTUK IDENTIFIKASI POTENSI SENYAWA BIOAKTIF PADA KEMUNING SEBAGAI ANTI-OBESITAS DENGAN *MACHINE LEARNING*

ZALANA ENDESNAH

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister 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.

Penguji Luar Komisi Pembimbing pada Ujian Tesis:
Dr. Eng Annisa, S.Kom, M.Kom.



Judul Tesis : Pengembangan Model *Drug-Target Interaction* untuk Identifikasi Potensi Senyawa Bioaktif pada Kemuning sebagai Anti-Obesitas dengan *Machine Learning*

Nama : Zalana Endesnah
NIM : G6501222044

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:
Prof. Dr. Eng. Wisnu Ananta Kusuma, S.T., M.T.

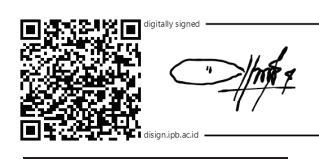


Pembimbing 2:
Dr. Mushthofa S.Kom., M.Sc.



Diketahui oleh

Ketua Program Studi:
Dr. Toto Haryanto, S.Kom., M.Si.
NIP 19821117 201404 1 001



Dekan Sekolah Sains Data, Matematika dan Informatika:
Prof. Dr. Ir Agus Buono, M.Si, M.Kom.
NIP. 19660702 199302 1 001



Tanggal Ujian:
31 Januari 2025

Tanggal Lulus:

PRAKATA

Puji syukur penulis panjatkan ke hadirat Allah Subhanahu wa Ta'ala atas segala rahmat, karunia, pertolongan, serta nikmat kesabaran, ketekunan, dan hidayah-Nya. Berkat limpahan kasih sayang-Nya yang menciptakan segala sesuatu dengan penuh hikmah, karya ilmiah ini akhirnya dapat diselesaikan dengan baik. Sungguh, hanya Allah tempat terbaik dalam meminta dan memohon pertolongan. Penelitian yang dilaksanakan sejak bulan April 2024 sampai bulan Januari 2025 ini penulis persembahkan untuk membuktikan tak satu pun yang Tuhan ciptakan tanpa manfaat bagi khalifah di muka bumi, yakni tanaman kemuning untuk pengobatan obesitas dengan judul “Pengembangan Model *Drug-Target Interaction* untuk Identifikasi Potensi Senyawa Bioaktif pada Kemuning sebagai Anti-obesitas dengan *Machine Learning*”.

Karya ilmiah ini tak mungkin selesai melalui tangan penulis seorang. Terdapat pembimbing yang senantiasa memberikan arahan, masukkan yang dapat menuntun penulis dalam pelaksanaan penelitian. Ucapan Terima kasih secara khusus penulis sampaikan kepada:

1. Bapak Prof. Dr. Eng. Wisnu Ananta Kusuma, S.T., M.T., Bapak Dr. Mushthofa, S.Kom., M.Sc., selaku komisi pembimbing yang telah membimbing, mengarahkan, dan memberikan nasehat terbaik kepada Penulis.
2. Ibu Dr. Dewi Anggraini Septaningsih, S.Si., M.Si. dan Mbak Sri Fauziah, A.Md., S.T., M.Si yang telah banyak membantu dan memberikan pendampingan kepada penulis dari proses pengumpulan data senyawa bioaktif kemuning sampai dengan tuntasnya karya ilmiah ini.
3. Kepada Ayahanda Drs. Zainal Arif, S.Sos, dan Ibunda Dra. Endel Tumur Juni M.Si, yang sangat penulis sayangi selalu memanjatkan doa, kasih sayang dan dukungan baik secara moral maupun material tiada hentinya, tanpa ridho dari beliau penulis tidak akan sampai pada titik ini.
4. Kepada teman-teman Ilmu Komputer angkatan 2022, teman-teman alumni *Puzzle Research Data Technology chapter* IPB University, serta pihak dan rekan lainnya yang tidak dapat penulis tuliskan satu per satu yang telah memberikan semangat, doa, dan dukungannya selama Penulis menempuh Pendidikan Magister di Bogor. Kota hujan yang dingin seolah menjadi hangat oleh hadirnya sahabat-sahabat dalam perjalanan ini. Semoga Allah SWT senantiasa melimpahkan rahmat dan karunia-Nya kepada kita semua.

Selesainya karya ilmiah ini penulis harapkan mampu memberikan kontribusi terhadap perkembangan ilmu pengetahuan terkhusus dalam upaya mencari kandidat senyawa obat obesitas yang berasal dari bahan alam yang melimpah seperti tanaman kemuning.

Bogor, Maret 2025

Zalana Endesnah



DAFTAR ISI

DAFTAR TABEL	ix
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	x
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Ruang Lingkup	3
TINJAUAN PUSTAKA	4
2.1 Senyawa Bioaktif	4
2.2 Kemuning (<i>Murraya Paniculata</i>)	4
2.3 Obesitas	5
2.4 <i>Machine Learning</i>	6
2.5 <i>Neighborhood Regularized Logistic Matrix Factorization (NRLMF)</i>	7
2.6 Random Forest	7
III METODE	9
3.1 Data Penelitian	9
3.2 Praproses Data	9
3.3 <i>Feature Engineering</i>	10
3.4 Pembangunan Matriks Input untuk NRLMF	12
3.5 NRLMF	13
3.6 Pemodelan DTI dengan <i>Machine Learning</i>	16
3.7 Evaluasi dan Seleksi Model	18
3.8 Prediksi DTI Senyawa Kemuning	19
3.9 Validasi Studi Literatur	20
3.10 Peralatan Penelitian	20
IV HASIL DAN PEMBAHASAN	21
4.1 Akuisisi Data dan Praproses	21
4.2 <i>Feature Engineering</i>	21
4.3 Pembangunan Matriks Input untuk NRLMF	22
4.4 Pembangkitan Interaksi Negatif menggunakan NRLMF	22
4.5 Pemodelan <i>Machine Learning</i> dan Evaluasi Model	22
4.6 Prediksi Interaksi	39
4.7 Validasi menggunakan Studi Literatur	40
V SIMPULAN DAN SARAN	45
5.1 Simpulan	45
5.2 Saran	45
DAFTAR PUSTAKA	46
LAMPIRAN	53
RIWAYAT HIDUP	62

DAFTAR TABEL

1	<i>Molecular fingerprint</i>	10
2	<i>Protein descriptor</i>	11
3	<i>Hyperparameter tuning</i>	17
4	Hasil kombinasi fitur senyawa-protein	21
5	Nilai <i>predicted probability</i> dan signifikansi interaksi terbaik hasil prediksi model RF.	39
6	Validasi hasil prediksi senyawa bioaktif kemuning	40

DAFTAR GAMBAR

1	Tumbuhan kemuning (<i>Murraya paniculata</i>)	4
2	Ilustrasi model RF	8
3	Tahapan penelitian	9
4	Ilustrasi ekstraksi fitur ECFP (Rogers dan Hahn 2010)	11
5	Ilustrasi <i>chemogenomics feature space</i> (Keterangan: r = jumlah baris, m = dimensi fitur protein, dan n = dimensi fitur senyawa)	11
6	Ilustrasi ruang matriks input untuk NRLMF	12
7	Persiapan data dan pembangunan matriks input untuk NRLMF	13
8	Matriks input algoritma NRLMF	15
9	Diagram alir untuk pembangkitan interaksi negatif menggunakan NRLMF.	16
10	Pemodelan 4 algoritma <i>machine learning</i> dengan 16 kombinasi fitur.	16
11	Ilustrasi alur kerja pemodelan	17
12	Grafik kurva ROC pada kombinasi fitur <i>AAIndex1_Morgan</i>	23
13	Diagram perbandingan akurasi setiap algoritma pada kombinasi fitur berbeda, tanda panah menandai model dengan performa terbaik dan kotak hitam bergaris putus-putus menandai kombinasi fitur terbaik.	25
14	Diagram perbandingan <i>precision</i> setiap algoritma pada kombinasi fitur berbeda, tanda panah menandai model dengan performa terbaik dan kotak hitam bergaris putus-putus menandai kombinasi fitur terbaik.	27
15	Diagram perbandingan <i>recall</i> setiap algoritma pada kombinasi fitur berbeda, tanda panah menandai model dengan performa terbaik dan kotak hitam bergaris putus-putus menandai kombinasi fitur terbaik.	29
16	Diagram perbandingan <i>F1-Score</i> setiap algoritma pada kombinasi fitur berbeda, tanda panah menandai model dengan performa terbaik dan kotak hitam bergaris putus-putus menandai kombinasi fitur terbaik.	31
17	Diagram perbandingan skor AUC setiap algoritma pada kombinasi fitur berbeda, tanda panah menandai model dengan performa terbaik dan kotak hitam bergaris putus-putus menandai kombinasi fitur terbaik.	32
18	Diagram perbandingan skor <i>cohen kappa</i> setiap algoritma pada kombinasi fitur berbeda, tanda panah menandai model dengan performa terbaik dan kotak hitam bergaris putus-putus menandai kombinasi fitur terbaik.	36



DAFTAR LAMPIRAN

1	Lampiran 1 Hasil matriks performa (Akurasi, Recall dan Precision) 64 model <i>machine learning</i> .	54
2	Lampiran 2 Hasil prediksi 68 pasangan interaksi senyawa kemuning dengan protein pada jaringan BAT.	56
3	Lampiran 3 Fungsi 20 protein target jaringan BAT hasil prediksi.	58

@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.