

**PENERAPAN METODE *RANDOM FOREST* UNTUK
PREDIKSI AKTIVITAS SENYAWA FLAVONOID DAN
TURUNANNYA DALAM MENGHAMBAT ENZIM α -
GLUKOSIDASE**

TUGAS AKHIR

Diajukan Oleh:

**PUTRI MAULAYA
NIM. 210705022**

**Mahasiswa Fakultas Sains dan Teknologi
Program Studi Teknologi Informasi**



**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS
ISLAM NEGERI AR-RANIRY BANDA ACEH
2025 M/1447 H**

LEMBAR PERSETUJUAN

PENERAPAN METODE *RANDOM FOREST* UNTUK PREDIKSI AKTIVITAS SENYAWA FLAVONOID DAN TURUNANNYA DALAM MENGHAMBAT ENZIM α - GLUKOSIDASE

TUGAS AKHIR

Diajukan kepada Fakultas Sains dan Teknologi UIN Ar-Raniry Banda Aceh
sebagai salah satu persyaratan penulisan Tugas Akhir dalam
Prodi Teknologi Informasi

Oleh:

Putri Maulaya
NIM. 210705022

Mahasiswa Fakultas Sains dan Teknologi
Program Studi Teknologi Informasi

Disetujui untuk Dimunaqasyahkan Oleh:

Pembimbing I,

Pembimbing II,

Dr. Hendri Ahmadian, S.Si., M.I.M Anjar Parba Asmara, M.Sc., Ph.D.
NIP. 198301042014031002 NIP. 198509092014031002

Mengetahui,

Ketua Program Studi Teknologi Informasi



Maishayati, M.T.
NIP.198301272015032003

LEMBAR PENGESAHAN

PENERAPAN METODE *RANDOM FOREST* UNTUK PREDIKSI AKTIVITAS SENYAWA FLAVONOID DAN TURUNANNYA DALAM MENGHAMBAT ENZIM α - GLUKOSIDASE

TUGAS AKHIR

Telah Diuji Oleh Panitia Ujian Munaqasyah Tugas Akhir
Fakultas Sains dan Teknologi UIN-Ar-Raniry Banda Aceh dan Dinyatakan Lulus
Serta Diterima Sebagai Salah Satu Beban Studi Program Sarjana (S-1)
Dalam Ilmu Teknologi Informasi

Pada Hari/Tanggal: Rabu, 2 Juli 2025 M
7 Muharram 1447 H
Di Darussalam, Banda Aceh

Panitia Ujian Munaqasyah Tugas Akhir:

Ketua,

Dr. Hendri Ahmadian, S.Si., M.I.M
NIP. 198301042014031002

Sekretaris,

Anjar Purba Asmara, M.Sc., Ph.D.
NIP. 198509092014031002

Pengaji I,

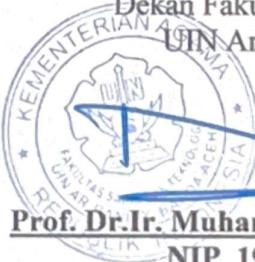
Khairan AR, M.Kom
NIP. 198607042014031001

Pengaji II,

Sarini Vita Dewi, S.T., M.Eng
NIP. 198712222022032001

Mengetahui:

Dekan Fakultas Sains dan Teknologi
UIN Ar-Raniry Banda Aceh,



Prof. Dr.Ir. Muhammad Dirhamsyah, M.T., I.P.U
NIP. 196210021988111001

LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertandatangan di bawah ini:

Nama : Putri Maulaya
NIM : 210705022
Program Studi : Teknologi Informasi
Fakultas : Sains dan Teknologi
Judul : Penerapan Metode *Random Forest* untuk Prediksi Aktivitas Senyawa Flavonoid dan Turunannya dalam Menghambat Enzim α -Glukosidase

Dengan ini menyatakan bahwa dalam penulisan tugas akhir ini, saya:

1. Tidak menggunakan ide orang lain tanpa mampu mengembangkan dan mempertanggungjawabkan;
2. Tidak melakukan plagiasi terhadap naskah karya orang lain;
3. Tidak menggunakan karya orang lain tanpa menyebutkan sumber asli atau tanpa izin pemilik karya;
4. Tidak memanipulasi dan memalsukan data;
5. Mengerjakan sendiri karya ini dan mampu bertanggungjawab atas karya ini.

Bila di kemudian hari ada tuntutan dari pihak lain atas karya saya, dan telah melalui pembuktian yang dapat dipertanggungjawabkan dan ternyata memang ditemukan bukti bahwa saya telah melanggar pernyataan ini, maka saya siap dikenai sanksi berdasarkan aturan yang berlaku di Fakultas Sains dan Teknologi UIN Ar-Raniry Banda Aceh.

Demikian pernyataan ini saya buat dengan sesungguhnya dan tanpa paksaan dari pihak manapun.

Banda Aceh, 15 Juli 2025

Yang Menyatakan,



Putri Maulaya

ABSTRAK

Nama	:	Putri Maulaya
Nim	:	210705022
Program Studi	:	Teknologi Informasi
Judul	:	Penerapan Metode <i>Random Forest</i> untuk Prediksi Aktivitas Senyawa Flavonoid dan Turunannya dalam Menghambat Enzim α -Glukosidase
Tanggal Sidang	:	2 Juli 2025
Jumlah Halaman	:	82 Halaman
Pembimbing I	:	Dr. Hendri Ahmadian, S.Si., M.I.M
Pembimbing II	:	Anjar Purba Asmara, M.Sc., Ph.D.
Kata Kunci	:	<i>Random Forest</i> , flavonoid, SMILES, Morgan Fingerprint, α -glukosidase, klasifikasi, machine learning.

Penelitian ini bertujuan untuk memprediksi aktivitas senyawa flavonoid dalam menghambat enzim α -glukosidase menggunakan algoritma *Random Forest* berbasis data struktur kimia. Representasi struktur senyawa diperoleh melalui dua cara, yaitu mengunduh file dalam format SDF dan mengekstrak manual struktur senyawa, yang kemudian dikonversi menjadi format SMILES. Data SMILES tersebut lalu diubah menjadi *Morgan Fingerprint* berdimensi 2048-bit sebagai fitur numerik. Dataset yang digunakan terdiri dari 100 senyawa yang seimbang, terdiri atas 50 senyawa aktif dan 50 tidak aktif. Model dilatih menggunakan parameter yang ditetapkan secara manual berdasarkan referensi terdahulu, tanpa melakukan *tuning* lebih lanjut. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasil pengujian menunjukkan akurasi prediksi sebesar 83%, dengan nilai presisi dan *recall* masing-masing sebesar 0,83. Analisis *confusion matrix* menunjukkan bahwa model memiliki kemampuan generalisasi yang baik dalam membedakan dua kelas aktivitas senyawa. Selain itu, hasil *feature importance* menunjukkan bahwa beberapa bit *fingerprint* memiliki kontribusi besar dalam proses klasifikasi. Dengan hasil tersebut, dapat disimpulkan bahwa algoritma *Random Forest* mampu mengklasifikasikan senyawa flavonoid berdasarkan struktur kimianya secara cukup efektif dan efisien. Penelitian ini menunjukkan potensi pemanfaatan metode pembelajaran mesin dalam bidang kimia komputasi, khususnya untuk mendukung skrining awal senyawa dari bahan alam yang berpotensi sebagai antidiabetes.

ABSTRACT

Name	:	Putri Maulaya
Student ID	:	210705022
Study Program	:	Information Technology
Title	:	<i>Application of the Random Forest Method for Predicting the Activity of Flavonoid Compounds and Their Derivatives in Inhibiting the α-Glucosidase Enzyme</i>
Thesis Defense Date	:	July 2, 2025
Total Pages	:	82 Pages
Supervisor I	:	Dr. Hendri Ahmadian, S.Si., M.I.M
Supervisor II	:	Anjar Purba Asmara, M.Sc., Ph.D.
Keywords	:	Random Forest, flavonoids, SMILES, Morgan Fingerprint, α -glucosidase, classification, machine learning.

This study aims to predict the activity of flavonoid compounds in inhibiting the α -glucosidase enzyme using the Random Forest algorithm based on chemical structure data. The molecular structures were obtained through two approaches: downloading files in SDF format and manually drawing molecular structures, which were then converted into SMILES format. The SMILES data were subsequently transformed into 2048-bit Morgan Fingerprint as numerical feature representations. The dataset used consisted of 100 balanced compounds, with 50 active and 50 inactive compounds. The model was trained using manually determined parameters based on prior research, without performing further hyperparameter tuning. The model evaluation was carried out using metrics such as accuracy, precision, recall, and F1-score. The results showed a prediction accuracy of 83%, with precision and recall values of 0.83 for each class. The confusion matrix analysis revealed that the model has good generalization ability in distinguishing between the two compound activity classes. Moreover, the feature importance analysis indicated that several fingerprint bits significantly contributed to the classification process. Based on these findings, it can be concluded that the Random Forest algorithm is capable of classifying flavonoid compounds based on their chemical structure effectively and efficiently. This study demonstrates the potential of machine learning-based methods in computational chemistry, particularly for supporting the early screening of natural compounds with potential as antidiabetic agents.

KATA PENGANTAR

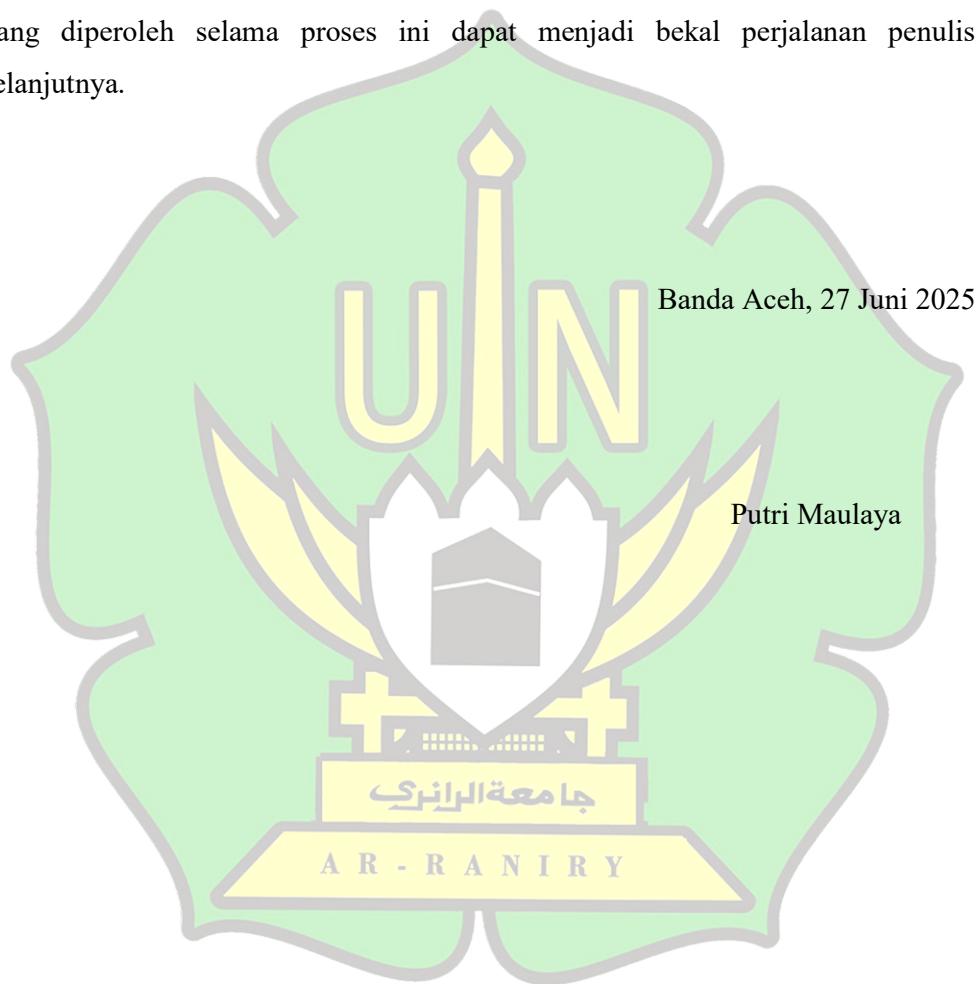
Puji syukur kepada Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan tugas akhir ini dengan judul “Penerapan Metode *Random Forest* untuk Prediksi Aktivitas Senyawa Flavonoid dan Turunannya dalam Menghambat Enzim α -Glukosidase”. Shalawat dan salam semoga senantiasa tercurahkan kepada junjungan Nabi Besar Muhammad SAW beserta keluarganya, para sahabatnya, dan seluruh umatnya yang selalu istiqamah hingga akhir zaman. Penulisan skripsi ini merupakan salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Ar-Raniry Banda Aceh.

Tugas akhir ini merupakan hasil dari proses panjang yang penuh perjuangan, kerja keras, tantangan, serta pembelajaran yang berharga. Dalam prosesnya, penulis menyadari bahwa pencapaian ini tidak akan terwujud tanpa doa, dukungan dan bimbingan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Kepada kedua orang tua tercinta yang selalu memberikan doa, dukungan, serta motivasi yang tiada henti selama penulis menjalani studi dan Kerja Praktik.
2. Bapak Dr. Hendri Ahmadian, S.Si., M.I.M dan Bapak Anjar Purba Asmara, M.Sc., Ph.D. selaku dosen pembimbing 1 dan dosen pembimbing 2 yang telah memberikan arahan serta masukan berharga kepada penulis dalam membuat tugas akhir ini.
3. Ibu Malahayati, M.T. selaku Ketua Program Studi Teknologi Informasi, yang telah membimbing, mengarahkan, dan memberikan ilmu yang sangat berharga kepada penulis.
4. Kepada Ibu Cut Ida Rahmadiana, S.Si., yang telah banyak membantu dalam berbagai urusan administrasi selama perkuliahan dan menyelesaikan tugas akhir ini.
5. Bapak/Ibu dosen dan seluruh staf pengajar di Program Studi Teknologi Informasi, yang telah membekali penulis dengan ilmu, nilai, dan wawasan selama masa perkuliahan.

6. Seluruh teman-teman seperjuangan di Prodi Teknologi Informasi, dosen penguji, serta pihak-pihak lainnya yang tidak dapat disebutkan satu per satu, namun telah memberikan bantuan dan kontribusi dalam bentuk apapun.

Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat dan menjadi kontribusi kecil dalam pengembangan ilmu pengetahuan, khususnya dalam bidang teknologi informasi dan kimia. Semoga segala ilmu dan pengalaman yang diperoleh selama proses ini dapat menjadi bekal perjalanan penulis selanjutnya.



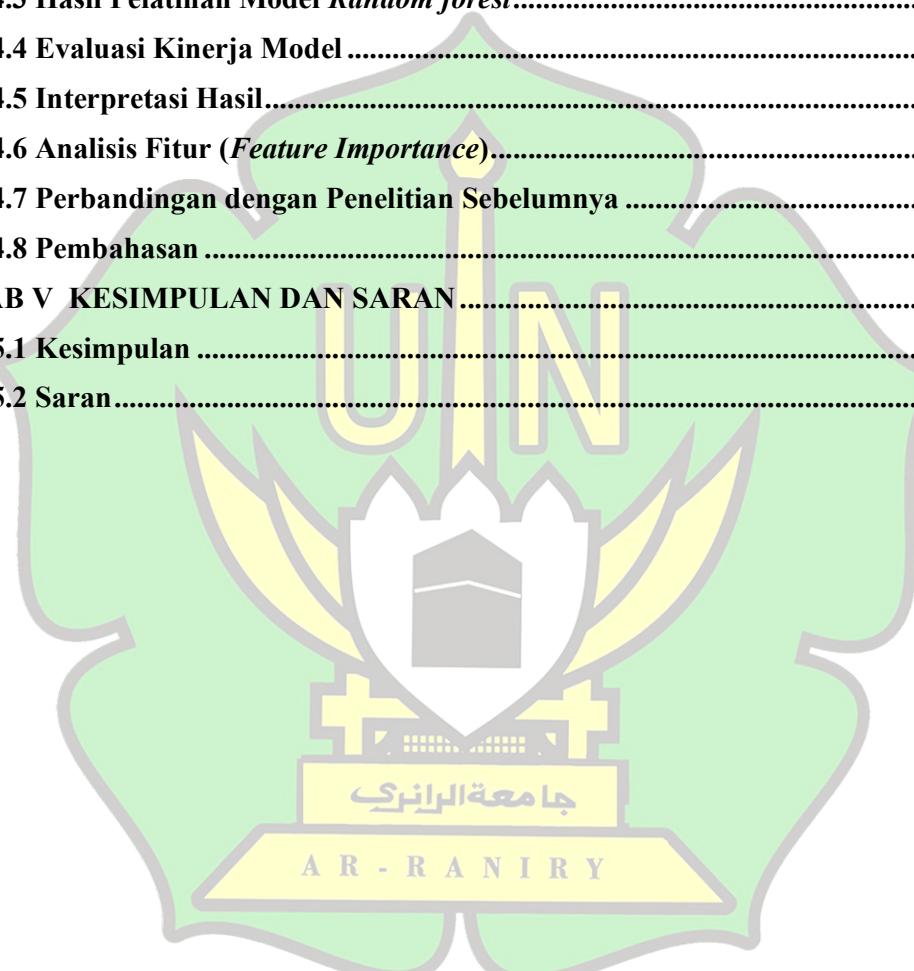
Banda Aceh, 27 Juni 2025

Putri Maulaya

DAFTAR ISI

LEMBAR PERSETUJUAN	i
LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR.....	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian	2
1.4 Manfaat Penelitian	2
1.5 Batasan Penelitian	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Penelitian Terdahulu.....	4
2.2 <i>Random forest</i>	7
2.3 Flavonoid dan Aktivitas Penghambatan Enzim α -Glukosidase.....	7
2.4 Enzim α -Glukosidase	8
2.5 Pengaplikasian <i>Random forest</i> dalam Kimia Komputasional	9
2.6 <i>Tuning Hyperparameter</i>	9
2.7 <i>Machine learning</i>	11
2.8 Data Kimia dan Representasi Molekul	11
2.9 Python.....	14
BAB III METODE PENELITIAN	16
3.1 Alat dan Perangkat Lunak	16
3.2 Metode Penelitian	16
3.3 Tahapan Penelitian.....	17
3.3.1 Pengumpulan Data.....	17
3.3.2 <i>Preprocessing Data</i>	27
3.3.3 Pemodelan dengan <i>Random forest</i>	27

3.3.4	Evaluasi Model	30
3.3.5	Tuning Hyperparameter	31
3.3.6	Hasilkan Prediksi	32
3.4	Waktu Penelitian	32
BAB IV HASIL DAN PEMBAHASAN		34
4.1	Deskripsi Dataset.....	34
4.2	Hasil <i>Preprocessing</i> Data.....	34
4.3	Hasil Pelatihan Model <i>Random forest</i>	37
4.4	Evaluasi Kinerja Model	38
4.5	Interpretasi Hasil.....	40
4.6	Analisis Fitur (<i>Feature Importance</i>).....	41
4.7	Perbandingan dengan Penelitian Sebelumnya	42
4.8	Pembahasan	43
BAB V KESIMPULAN DAN SARAN		46
5.1	Kesimpulan	46
5.2	Saran	46



DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian	17
Gambar 3. 2 Arsitektur Random forest (Chen dkk., 2019).....	28
Gambar 4. 1 Proses validasi struktur SMILES dengan output jumlah senyawa valid dan tidak valid	35
Gambar 4. 2 Kode Python untuk mengonversi SMILES menjadi Morgan Fingerprint berdimensi 2048-bit	36
Gambar 4. 3 Struktur kolom file flavonoids_with_fingerprint.csv hasil konversi fingerprint	36
Gambar 4. 4 Cuplikan isi data fingerprint untuk senyawa Galangin dan Baicalein	37
Gambar 4. 5 Confusion Matrix Hasil Klasifikasi Model Random forest	39
Gambar 4. 6 Feature Importance.....	42



DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	5
Tabel 3. 1 Spesifikasi Perangkat Keras.....	16
Tabel 3. 2 Spesifikasi Perangkat Lunak.....	16
Tabel 3. 3 Dataset yang digunakan	19
Tabel 3. 4 Timeline penggerjaan.....	32
Tabel 4. 1 Parameter Random Forest.....	37
Tabel 4. 2 Hasil evaluasi metrik.....	38
Tabel 4. 3 Hasil Prediksi Model terhadap Data Uji	40
Tabel Lampiran 5. 1 Confusion Matrix Model Random Forest.....	70



BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi telah membawa dampak signifikan dalam berbagai bidang, termasuk kimia komputasi. Integrasi antara teknologi informasi dan kimia memungkinkan analisis data yang lebih efisien dan akurat, terutama dalam memprediksi aktivitas senyawa kimia. Salah satu pendekatan yang digunakan adalah penerapan algoritma *machine learning* untuk memprediksi aktivitas senyawa tertentu terhadap target biologis (Jamir & Hariprasad, 2024).

Di sisi lain, dalam bidang kimia, banyak penelitian yang berfokus pada eksplorasi senyawa alami yang memiliki potensi aktivitas biologis tertentu. Salah satu kelompok senyawa yang menarik perhatian adalah flavonoid, yang diketahui memiliki berbagai manfaat kesehatan, termasuk sebagai inhibitor enzim α -glukosidase. Enzim ini berperan penting dalam metabolisme karbohidrat, sehingga penghambatannya dapat menjadi strategi potensial dalam pengembangan obat antidiabetes. Namun, penentu efektivitas flavonoid sebagai inhibitor enzim ini masih membutuhkan serangkaian uji laboratorium yang kompleks dan memakan waktu. Oleh karena itu, diperlukan pendekatan alternatif yang lebih efisien untuk memprediksi aktivitas senyawa sebelum diuji secara eksperimental (Puspita et al., 2023).

Machine learning menawarkan solusi dalam menganalisis dan memprediksi aktivitas flavonoid berdasarkan data strukturnya. Dengan menggunakan teknik pembelajaran mesin, model prediktif dapat dikembangkan untuk memperkirakan seberapa efektif suatu senyawa flavonoid dalam menghambat enzim α -glukosidase hanya berdasarkan strukturnya. Pendekatan ini tidak hanya menghemat waktu dan biaya penelitian, tetapi juga mempercepat proses skrining senyawa potensial yang dapat dikembangkan lebih lanjut dalam penelitian kimia (Jamir & Hariprasad, 2024). Sejauh ini, penelitian tentang analisis senyawa flavonoid sebagai inhibitor enzim ini dengan *machine learning* masih sangat terbatas.

Dalam penelitian ini, digunakan algoritma *random forest* untuk membangun model prediksi aktivitas flavonoid. *Random forest* dipilih karena merupakan metode yang *robust*, mampu menangani data dengan banyak fitur, serta mengurangi risiko *overfitting* dibandingkan dengan model *decision tree* tunggal. Algoritma ini

bekerja dengan membangun banyak pohon keputusan (*decision trees*) dan menggabungkan hasilnya untuk menghasilkan prediksi yang lebih akurat. Selain itu, *random forest* mampu memberikan interpretasi fitur-fitur yang paling berpengaruh dalam menentukan aktivitas senyawa, yang dapat memberikan wawasan lebih lanjut dalam penelitian kimia komputasi (Wuryani & Agustiani, 2021).

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan *machine learning*, khususnya algoritma *random forest*, dalam memprediksi aktivitas flavonoid sebagai inhibitor enzim α -glukosidase. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode prediktif berbasis teknologi informasi yang dapat mendukung penelitian kimia dalam eksplorasi bioaktif.

1.2 Rumusan Masalah

Adapun rumusan masalah yang diangkat pada penulisan ini adalah sebagai berikut:

1. Bagaimana penerapan metode *random forest* dalam memprediksi aktivitas senyawa flavonoid dan turunannya dalam menghambat enzim α -glukosidase?
2. Bagaimana kinerja metode *random forest* dalam memprediksi aktivitas senyawa flavonoid berdasarkan metrik evaluasi seperti akurasi, presisi, dan *recall*?

1.3 Tujuan Penelitian

Adapun tujuan penulisan ini adalah sebagai berikut:

1. Menerapkan metode *random forest* untuk memprediksi aktivitas senyawa flavonoid dalam menghambat enzim α -glukosidase.
2. Mengevaluasi kinerja model *random forest* dalam konteks kimia komputasional menggunakan metrik evaluasi seperti akurasi, presisi, dan *recall*.

1.4 Manfaat Penelitian

Berikut adalah manfaat yang diharapkan dari penulisan skripsi:

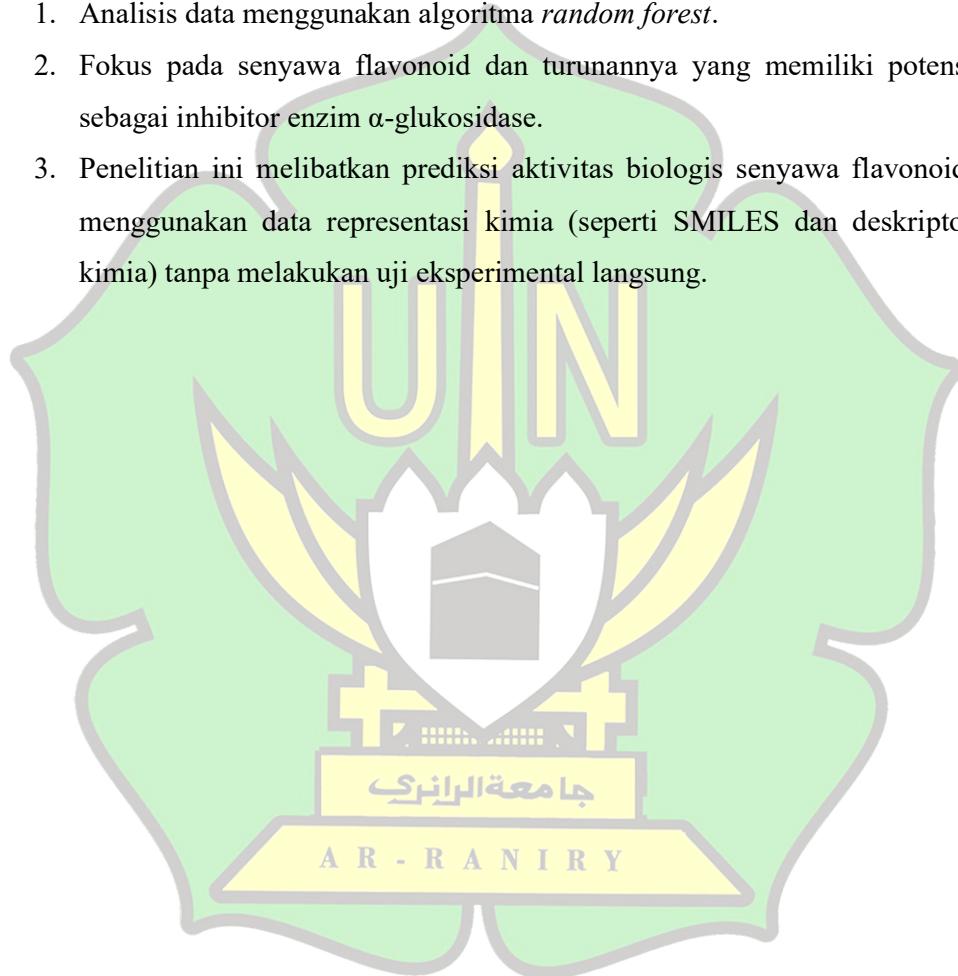
1. Memberikan kontribusi terhadap perkembangan teknik *machine learning* dalam bidang kimia komputasional dan bioinformatika.

2. Memperluas penerapan metode *random forest* dalam prediksi aktivitas biologis senyawa kimia.
3. Mendukung penelitian obat antidiabetes melalui identifikasi senyawa potensial sebagai inhibitor enzim α -glukosidase.

1.5 Batasan Penelitian

Adapun ruang lingkup pada penelitian ini adalah sebagai berikut:

1. Analisis data menggunakan algoritma *random forest*.
2. Fokus pada senyawa flavonoid dan turunannya yang memiliki potensi sebagai inhibitor enzim α -glukosidase.
3. Penelitian ini melibatkan prediksi aktivitas biologis senyawa flavonoid, menggunakan data representasi kimia (seperti SMILES dan deskriptor kimia) tanpa melakukan uji eksperimental langsung.



BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Menurut Saranani dkk. (2023), daun Kirinyuh (*Chromolaena odorata* L.) memiliki potensi sebagai agen antidiabetes alami karena kandungan senyawa aktif seperti flavonoid, tanin, saponin, dan terpenoid yang berperan dalam menurunkan kadar glukosa darah. Penelitian tersebut mengkaji aktivitas penghambatan enzim α -glukosidase secara *in vitro* melalui ekstraksi etanol daun kirinyuh menggunakan metode maserasi dengan pelarut etanol 96%. Uji dilakukan pada empat tingkat konsentrasi ekstrak (25%, 50%, 100%, dan 200%) berdasarkan metode yang diadaptasi dari Salehi. Hasil pengujian menunjukkan nilai konsentrasi daya hambat 50% (IC_{50}) sebesar 65,59 ppm pada konsentrasi optimal, yang mengindikasikan efektivitas penghambatan terhadap enzim target.

Amaliah dkk. (2022) meneliti penggunaan metode *Random forest* (RF) dalam meningkatkan akurasi prediksi varian minuman kopi yang paling diminati pelanggan di kedai Konijiwa Bantaeng. Model dengan parameter mtry 2 dan ntree 500 menunjukkan tingkat kesalahan klasifikasi terendah dan dievaluasi menggunakan *confusion matrix*. Hasil analisis menunjukkan bahwa varian *coffee-based* lebih diminati dibandingkan *signature coffee*, dengan akurasi mencapai 94,12%.

Meiriyama dan Sudiadi (2022) menyoroti potensi tanaman obat di Indonesia dalam pengobatan tradisional. Namun, pemahaman masyarakat mengenai tanaman herbal masih terbatas. Oleh karena itu, penelitian ini menggunakan algoritma *Random forest* dengan fitur *Histogram of Oriented Gradients* (HOG) untuk mengklasifikasikan daun herbal. Citra daun diubah menjadi format *greyscale* dengan resolusi 816×612 piksel sebelum diekstraksi menggunakan metode HOG, menghasilkan vektor fitur berukuran 1×3168 . Model *Random forest* yang dikembangkan mencapai akurasi 85,33% dalam mengklasifikasikan jenis daun herbal.

Nasukha dkk. (2023) meneliti potensi daun salam sebagai inhibitor enzim α -glukosidase untuk terapi diabetes. Ekstrak etil asetat dan fraksi hasil kromatografi kolom diuji menggunakan *ELISA reader* pada panjang gelombang 405 nm. Hasil penelitian menunjukkan bahwa ekstrak etil asetat memiliki nilai IC_{50} sebesar 87,93

ppm, sedangkan fraksi terbaik (F-5,5) memiliki nilai IC₅₀ sebesar 18,99 ppm, yang mendekati efektivitas inhibitor standar, akarbosa (18,05 ppm).

Jamir dan Hariprasad (2024) mengembangkan model *machine learning* (ML) untuk memprediksi potensi inhibitor α -glukosidase dari 537 metabolit sekunder tanaman (plant secondary metabolites / PSM). Dari 35 model ML yang diuji, model *Random forest* berbasis sidik jari RDKit menunjukkan performa terbaik dengan akurasi 83,74% dan *area under the curve* (AUC) sebesar 0,803. Model ini diterapkan pada 5810 senyawa dalam basis data PSM, menghasilkan 965 kandidat dengan probabilitas prediksi $\geq 0,90$. Dari 24 senyawa yang kemudian diuji secara *in vitro*, sebanyak 13 senyawa menunjukkan aktivitas penghambatan dengan nilai IC₅₀ antara 0,63 hingga 7 mg/mL, di mana tujuh di antaranya lebih efektif dibandingkan akarbosa. Asam nervonat diidentifikasi sebagai kandidat potensial untuk pengembangan terapi diabetes lebih lanjut.

Proenca dkk. (2017) meneliti potensi flavonoid sebagai inhibitor α -glukosidase dalam menekan hiperglikemia postprandial (*postprandial hyperglycemia* / PPHG). Sebanyak 44 flavonoid dari lima kelompok dianalisis berdasarkan studi hubungan struktur-aktivitas (*structure-activity relationship* / SAR) secara *in vitro*. Hasil penelitian menunjukkan bahwa senyawa dengan gugus katekol pada cincin A dan B, serta gugus 3-OH pada cincin C, memiliki aktivitas penghambatan tertinggi dengan nilai IC₅₀ yang lebih rendah dibandingkan akarbosa. Studi ini menunjukkan bahwa beberapa flavonoid berpotensi sebagai alternatif dalam pengelolaan PPHG.

Tabel 2. 1 Penelitian Terdahulu

No.	Referensi	Metode	Hasil
1.	Saranani et al., (2023)	Ekstraksi menggunakan pelarut etanol, diikuti dengan uji aktivitas penghambatan secara <i>in vitro</i> .	Ekstrak menunjukkan aktivitas penghambatan enzim α -glukosidase, diduga karena kandungan saponin dan flavonoid.
2.	Amaliah et al., (2022)	<i>Random forest</i> digunakan untuk mengklasifikasi data	<i>Random forest</i> mencapai akurasi sebesar 94,12%, menunjukkan

No.	Referensi	Metode	Hasil
		berdasarkan variabel input dengan parameter mtry dan ntree.	kemampuannya untuk menangani data klasifikasi.
3.	Meiriyama & Sudiadi, (2022)	<i>Random forest</i> digunakan untuk mengklasifikasikan jenis daun herbal berdasarkan fitur <i>Histogram of Oriented Gradients (HOG)</i> .	Model <i>random forest</i> mencapai akurasi sebesar 85,33% dalam mengklasifikasi jenis daun herbal.
4.	Nasukha et al., (2023)	Ekstraksi menggunakan etil asetat dan fraksinasi menggunakan kromatografi kolom, diikuti dengan uji <i>in vitro</i> .	Fraksi tertentu menunjukkan aktivitas penghambatan yang signifikan, diduga karena kandungan flavonoid.
5.	Jamir & Hariprasad, (2024)	Membangun 35 model <i>machine learning</i> , termasuk <i>Random forest</i> , untuk memprediksi potensi inhibitor α -glukosidase dari metabolit sekunder tanaman.	Model <i>Random forest</i> dengan sidik jari RDKit mencapai akurasi 83,74% dan AUC 0,803, mengidentifikasi 965 senyawa dengan potensi penghambatan α -glukosidase.
6.	Proença et al., (2017)	Studi eksperimental dan simulasi <i>in silico</i> untuk menganalisis aktivitas penghambatan flavonoid terhadap enzim α -glukosidase.	Beberapa flavonoid seperti quercetin dan luteolin menunjukkan efek penghambatan kuat terhadap α -glukosidase dengan IC ₅₀ lebih rendah dibandingkan akarbosa.

2.2 Random forest

Random forest (RF) adalah algoritma *machine learning* berbasis *ensemble* yang digunakan untuk membangun model prediksi dalam berbagai konteks penelitian. Algoritma ini terdiri dari sejumlah besar pohon keputusan yang dibuat secara acak dari subset data pelatihan dan variabel prediktor. Salah satu keunggulan utama RF adalah kemampuannya menangani dataset dengan jumlah variabel yang besar, sekaligus mengurangi risiko *overfitting* melalui mekanisme agregasi hasil dari beberapa pohon keputusan.

Dalam penerapannya, RF sering digunakan untuk mengembangkan model klasifikasi dan regresi. Setiap pohon dalam RF memberikan prediksi sendiri, dan hasil akhir diperoleh melalui voting mayoritas (untuk klasifikasi) atau rata-rata prediksi (untuk regresi). RF juga memungkinkan pemilihan variabel penting dengan mengukur kontribusi setiap variabel terhadap akurasi prediksi. Oleh karena itu, metode ini sering digunakan dalam penelitian biomedis, bioinformatika, dan pengolahan data berskala besar.

Meskipun RF menawarkan akurasi tinggi dan stabilitas dalam prediksi, salah satu tantangannya adalah efisiensi komputasi, terutama saat menangani dataset yang sangat besar. Untuk mengatasi hal ini, beberapa metode pemilihan variabel telah dikembangkan guna menyederhanakan model tanpa mengorbankan akurasi secara signifikan (Speiser et al., 2019).

2.3 Flavonoid dan Aktivitas Penghambatan Enzim α -Glukosidase

Flavonoid merupakan senyawa polifenol alami yang banyak ditemukan dalam tanaman dan memiliki berbagai aktivitas biologis, termasuk sebagai antioksidan dan agen terapeutik potensial dalam pengelolaan penyakit metabolismik seperti diabetes mellitus tipe 2. Salah satu mekanisme utama dalam mengendalikan kadar glukosa darah adalah dengan menghambat enzim α -glukosidase, yang berperan dalam pemecahan karbohidrat kompleks menjadi gula sederhana di usus.

Inhibitor α -glukosidase dapat memperlambat penyerapan glukosa ke dalam darah, sehingga mengurangi lonjakan kadar glukosa pasca makan. Flavonoid diketahui memiliki kemampuan untuk menghambat aktivitas enzim ini melalui interaksi dengan situs aktif enzim, yang ditentukan oleh struktur kimia mereka. Penelitian menunjukkan bahwa flavonoid dengan gugus hidroksil tertentu dan

struktur spesifik pada cincin aromatiknya memiliki potensi penghambatan yang lebih tinggi dibandingkan inhibitor konvensional seperti akarbosa.

Studi *in vitro* dan *in silico* telah dilakukan untuk mengidentifikasi efektivitas flavonoid dalam menghambat α -glukosidase. Analisis kinetik menunjukkan bahwa beberapa flavonoid berperan sebagai inhibitor kompetitif atau non-kompetitif, tergantung pada struktur dan posisi gugus fungsionalnya. Molekul flavonoid seperti quercetin dan kaempferol telah ditemukan memiliki aktivitas penghambatan yang lebih baik dibandingkan beberapa obat sintetik, dengan nilai IC₅₀ yang lebih rendah, menunjukkan potensi penggunaannya dalam terapi diabetes (Proença et al., 2017).

Penelitian oleh Proenca dkk. (2017) menguji 44 jenis flavonoid dan menemukan bahwa beberapa senyawa memiliki aktivitas penghambatan α -glukosidase yang signifikan. Hasil ini menunjukkan bahwa struktur kimia flavonoid, seperti keberadaan gugus hidroksil pada posisi tertentu, berkontribusi terhadap potensi penghambatan enzim. Dengan menghambat aktivitas enzim ini, penyerapan glukosa dapat diperlambat, sehingga membantu mengontrol kadar gula darah postprandial pada penderita diabetes.

Melalui pendekatan *docking* molekuler, interaksi antara flavonoid dan residu aktif enzim α -glukosidase dapat dipelajari lebih lanjut untuk mengembangkan senyawa yang lebih efektif sebagai agen antidiabetes. Oleh karena itu, flavonoid menjadi fokus penelitian sebagai alternatif alami dalam pengelolaan hiperglikemia dan dapat dikembangkan lebih lanjut sebagai kandidat obat yang lebih aman dengan efek samping yang lebih minimal.

2.4 Enzim α -Glukosidase

α -Glukosidase adalah enzim yang berperan dalam proses metabolisme karbohidrat dengan memecah oligosakarida dan polisakarida menjadi monosakarida di usus halus manusia. Proses ini sangat penting dalam pencernaan dan penyerapan glukosa ke dalam darah. Senyawa yang mampu menghambat aktivitas enzim ini berpotensi digunakan sebagai agen antidiabetes karena dapat memperlambat penyerapan glukosa dan menurunkan kadar gula darah setelah makan.

Salah satu inhibitor α -glukosidase yang umum digunakan adalah akarbose, yang bekerja dengan menghambat enzim sehingga tidak semua karbohidrat kompleks dapat dipecah menjadi glukosa. Dengan demikian, penggunaan inhibitor

α -glukosidase dapat membantu mencegah lonjakan kadar gula darah setelah makan dan mengurangi risiko komplikasi diabetes mellitus tipe 2 (Sari et al., 2020).

Studi sebelumnya oleh Puspita dkk. (2023) menunjukkan bahwa flavonoid dari daun kelor (*Moringa oleifera*) memiliki potensi sebagai inhibitor alami enzim α -glukosidase, baik melalui pendekatan *in silico* maupun *in vitro*. Hal ini mengindikasikan bahwa senyawa alami dapat menjadi sumber potensial untuk pengembangan obat antidiabetes yang lebih aman dan efektif.

2.5 Pengaplikasian *Random forest* dalam Kimia Komputasional

Random forest (RF) merupakan salah satu algoritma pembelajaran mesin yang banyak digunakan dalam analisis kimia komputasional, terutama dalam bidang penemuan obat berbantuan komputer (CADD). Algoritma ini bekerja dengan membangun banyak pohon keputusan secara acak dan menggabungkan hasilnya untuk meningkatkan akurasi prediksi, menjadikannya metode yang andal dalam pemrosesan data berdimensi tinggi dengan sampel terbatas tanpa menyebabkan *overfitting* (Jamir & Hariprasad, 2024).

Dalam bidang kimia komputasional, *random forest* digunakan untuk memprediksi aktivitas biologis senyawa kimia berdasarkan data representasi molekul, seperti SMILES atau deskriptor kimia. Algoritma ini tidak hanya mampu menangani dataset yang besar, tetapi juga memberikan wawasan tentang variabel-variabel yang paling berpengaruh terhadap aktivitas biologis senyawa.

Penelitian oleh Jamir dan Hariprasad (2024) menunjukkan bahwa *random forest* dapat digunakan untuk memprediksi senyawa bioaktif yang berpotensi sebagai inhibitor α -glukosidase. Dengan menggunakan data sekunder metabolit tanaman, algoritma ini mampu mengidentifikasi senyawa dengan aktivitas biologis tinggi, yang kemudian divalidasi melalui uji *in vitro*. Pendekatan ini tidak hanya mempercepat proses penemuan senyawa baru, tetapi juga mengurangi biaya penelitian.

Dengan kemampuannya dalam menangani dataset besar dan kompleks, RF terus berkembang sebagai alat penting dalam kimia komputasional, terutama dalam identifikasi dan desain molekul baru yang berpotensi sebagai agen terapeutik.

2.6 Tuning Hyperparameter

Tuning hyperparameter merupakan salah satu langkah penting dalam proses pelatihan model *machine learning*. *Hyperparameter* adalah parameter yang

nilainya ditentukan sebelum proses pelatihan dimulai, dan memiliki pengaruh signifikan terhadap performa akhir model. Pada algoritma *random forest*, *hyperparameter* yang umum disesuaikan antara lain jumlah pohon (*n_estimators*), kedalaman maksimum pohon (*max_depth*), serta kriteria pemisahan node (*criterion*). Tujuan utama dari *tuning* adalah untuk menemukan kombinasi nilai yang memberikan hasil prediksi terbaik terhadap data uji.

Salah satu teknik yang sering digunakan dalam *tuning hyperparameter* adalah *Grid Search*, yaitu metode pencarian sistematis yang menguji semua kemungkinan kombinasi *hyperparameter* yang telah ditentukan sebelumnya. Proses ini biasanya disertai dengan validasi silang (*cross-validation*) agar model yang dihasilkan tidak hanya akurat terhadap data latih, tetapi juga memiliki kemampuan generalisasi terhadap data baru. Dalam konteks penelitian ini, *tuning* direncanakan sebagai bagian dari eksperimen untuk melihat apakah performa model dapat ditingkatkan lebih lanjut melalui penyetelan *hyperparameter* tersebut (Weerts et al., 2020).

Namun demikian, *tuning* tidak selalu diperlukan, terutama apabila model yang dibangun dengan parameter default telah menunjukkan hasil yang baik. Misalnya, jika model mampu mencapai akurasi sebesar 83% tanpa *tuning*, maka performa tersebut sudah tergolong optimal dalam banyak studi klasifikasi senyawa kimia. Weerts et al. (2020) menunjukkan bahwa dalam beberapa kasus, penggunaan *hyperparameter* default pada *Random Forest* justru menghasilkan performa yang sebanding, atau bahkan lebih baik daripada *tuning* dengan metode pencarian acak terbatas. Oleh karena itu, *tuning* sebaiknya dilakukan hanya bila akurasi awal belum memuaskan.

Dalam penelitian ini, *tuning* disiapkan sebagai salah satu tahap yang akan dilakukan apabila akurasi awal model belum mencapai tingkat yang diharapkan. Akan tetapi, jika akurasi model tanpa *tuning* sudah menunjukkan performa yang tinggi dan stabil, maka *tuning* dapat dipertimbangkan untuk tidak dilaksanakan demi menjaga efisiensi waktu dan menghindari kompleksitas model yang tidak perlu. Dengan pendekatan ini, proses *tuning* tetap disertakan dalam desain metodologi, namun pelaksanaannya bersifat adaptif tergantung pada hasil evaluasi awal.

2.7 Machine learning

Machine learning (ML) adalah cabang dari kecerdasan buatan (*Artificial Intelligence/AI*) yang memungkinkan sistem komputer belajar dari data tanpa diprogram secara eksplisit. Teknik ini digunakan untuk mengidentifikasi pola dalam data dan membuat prediksi berdasarkan pola tersebut (Yakut, 2023). Dalam konteks penelitian ini, ML digunakan untuk memprediksi aktivitas senyawa flavonoid dalam menghambat enzim α -glukosidase, yang berperan dalam metabolisme karbohidrat dan memiliki relevansi dalam pengobatan diabetes.

ML telah digunakan dalam berbagai bidang, termasuk prakiraan cuaca, pengenalan gambar, pengolahan bahasa alami, hingga prediksi hasil medis. Dalam ranah biomedis, ML sering digunakan untuk membangun model prediktif yang dapat mengidentifikasi hubungan kompleks antara variabel klinis. Beberapa algoritma ML yang umum digunakan meliputi Jaringan Syaraf Tiruan (*Artificial Neural Network/ANN*), Mesin Vektor Pendukung (*Support Vector Machine/SVM*), Pohon Keputusan (*Decision Tree/DT*), dan Hutan Acak (*Random forest/RF*).

Meskipun ML memiliki banyak keunggulan, penerapannya dalam dataset kecil masih menjadi tantangan, karena jumlah sampel yang terbatas dapat menyebabkan model menjadi kurang stabil dan rentan terhadap *overfitting*. Oleh karena itu, pemilihan algoritma yang tepat serta teknik pengolahan data yang baik sangat penting untuk meningkatkan akurasi prediksi dalam penelitian biomedis (Shaikhina et al., 2019).

Agar model ML bekerja secara optimal, diperlukan proses validasi dan evaluasi menggunakan teknik seperti *k-fold cross-validation* untuk memastikan model tidak *overfitting*. Selain itu, metrik evaluasi seperti akurasi, presisi, *recall*, *F1-score*, dan *Area Under Curve* (AUC) digunakan untuk menilai performa model. Dengan pendekatan ini, ML dapat digunakan dalam analisis data kimia untuk mengidentifikasi senyawa yang berpotensi sebagai inhibitor enzim, sehingga mendukung pengembangan terapi berbasis flavonoid bagi penderita diabetes (Yakut, 2023).

2.8 Data Kimia dan Representasi Molekul

Data Kimia dan Representasi Molekul membahas bagaimana molekul direpresentasikan dalam bentuk data numerik atau simbolik untuk keperluan analisis, prediksi sifat, dan desain molekul baru. Kemajuan dalam ilmu kimia dan

kecerdasan buatan (AI) telah menghasilkan berbagai pendekatan representasi molekul yang lebih akurat dan informatif, seperti sistem notasi kimia SMILES, pemrosesan dan ekstraksi fitur menggunakan pustaka kimia RDKit, serta teknik *fingerprinting* seperti Morgan *Fingerprint*. Pendekatan-pendekatan ini memiliki peran penting dalam mengubah struktur kimia menjadi bentuk yang dapat dipahami dan dianalisis secara komputasional, yang masing-masing akan dijelaskan lebih lanjut pada bagian berikut:

A. SMILES (*Simplified Molecular Input Line Entry System*)

SMILES merupakan representasi linear dari struktur molekul kimia yang banyak digunakan dalam basis data kimia modern seperti ChEMBL. Dalam konteks kimia komputasional, SMILES memiliki keunggulan karena dapat dikonversi langsung dari dan ke bentuk struktur molekul secara efisien. Format ini memungkinkan molekul digambarkan sebagai urutan karakter yang mencerminkan atom dan ikatan secara sintaksis. Oleh karena bentuknya yang seragam dan sederhana, SMILES telah menjadi format standar untuk masukan (*input*) dalam berbagai model *machine learning* dan *deep learning*.

Setiap molekul diubah menjadi *string* karakter berdasarkan aturan tertentu. Misalnya, atom karbon ditulis sebagai "C", oksigen sebagai "O", ikatan rangkap ditulis dengan simbol "=?", dan struktur bercabang ditandai dengan tanda kurung. Sebagai contoh, molekul etanol ditulis sebagai CCO dalam format SMILES, yang menunjukkan dua atom karbon dan satu oksigen terhubung secara linear.

Dalam penelitian Honda et al. (2019), SMILES digunakan sebagai dasar input untuk pelatihan model Transformer, yaitu arsitektur deep *learning* berbasis urutan (*sequence*) yang sebelumnya banyak digunakan dalam pemrosesan bahasa alami (NLP). Model tersebut dilatih menggunakan lebih dari 860.000 data SMILES untuk menghasilkan *fingerprint* molekul berdimensi 1024 yang bersifat data-driven. Pendekatan ini memungkinkan representasi molekul tidak lagi bergantung pada *fingerprint* tradisional seperti ECFP, melainkan dibentuk secara otomatis oleh pemahaman model terhadap pola-pola kimia yang terkandung dalam urutan SMILES itu sendiri.

Penelitian ini menunjukkan bahwa SMILES bukan hanya representasi struktural, tetapi juga dapat dieksplorasi sebagai bentuk representasi semantik dari molekul. Dengan pelatihan yang cukup, model Transformer mampu

mengekstrak informasi kontekstual dari SMILES secara mendalam, sehingga *fingerprint* yang dihasilkan lebih informatif untuk tugas seperti prediksi bioaktivitas atau klasifikasi senyawa. Hasil ini menegaskan bahwa SMILES tidak hanya berguna sebagai input untuk sistem *rule-based*, tetapi juga memiliki potensi besar dalam pengembangan model prediktif berbasis *deep learning* (Honda et al., 2019).

B. RDKit

RDKit adalah pustaka (*library*) *open-source* yang digunakan untuk komputasi kimia. Dalam konteks penelitian ini, RDKit dimanfaatkan untuk mengekstraksi fitur molekul dari data SMILES. SMILES adalah representasi tekstual dari struktur kimia suatu senyawa, dan RDKit dapat memproses representasi ini untuk menghasilkan berbagai macam informasi dan fitur kimia yang dapat digunakan sebagai input dalam proses pemodelan *machine learning*.

Fungsi RDKit dalam penelitian ini meliputi:

- Mengkonversi data SMILES menjadi representasi molekul.
- Mengekstraksi berbagai deskriptor molekuler seperti berat molekul, jumlah atom tertentu, polaritas, dan lain-lain.
- Memberikan representasi numerik yang diperlukan untuk pemodelan menggunakan algoritma seperti *Random forest*.

RDKit sangat berguna dalam *cheminformatics* karena kemampuannya yang luas untuk mengolah dan menganalisis data kimia secara efisien, serta dukungannya terhadap berbagai operasi struktur kimia seperti perhitungan deskriptor, pencocokan substruktur, dan konversi antar format (BYADI et al., 2025).

C. Morgan Fingerprints

Morgan *fingerprint* merupakan salah satu metode representasi molekul secara digital yang banyak digunakan dalam bidang kimia komputasi dan *cheminformatics*. *Fingerprint* ini dikembangkan berdasarkan algoritma Morgan dan menjadi dasar dari *Extended-Connectivity Fingerprints* (ECFP), seperti ECFP4 dan ECFP6. *Fingerprint* ini memungkinkan deskripsi struktur molekul dalam bentuk bitstring berdimensi tetap yang dapat digunakan sebagai input dalam proses pembelajaran mesin (*machine learning*) atau pencarian kemiripan struktur kimia.

Morgan *fingerprint* bekerja dengan cara mengevaluasi lingkungan atom dalam molekul berdasarkan radius tertentu. Radius ini menentukan sejauh mana atom-atom di sekitar atom pusat akan dipertimbangkan. Sebagai contoh, radius 2 akan mencakup atom pusat dan semua atom yang terhubung hingga dua ikatan. Fragmen molekul yang diperoleh dari proses ini kemudian diubah menjadi representasi numerik melalui proses hashing, dan hasilnya disimpan dalam bentuk vektor biner berdimensi tetap, misalnya 1024 bit.

Dalam penelitian ini, Morgan *fingerprint* dihasilkan menggunakan pustaka RDKit dengan parameter radius 2 dan panjang bit 1024. Hal ini setara dengan ECFP4 dalam literatur *cheminformatics*. Vektor biner hasil representasi *fingerprint* tersebut digunakan sebagai fitur input dalam proses pemodelan *machine learning* untuk memprediksi sifat atau aktivitas molekul.(LibreTexts, 2025)

Implementasi Morgan *fingerprint* sangat bermanfaat karena kemampuannya dalam merepresentasikan informasi lokal dari struktur molekul secara efisien, serta kompatibilitasnya dengan berbagai metode pemodelan data.

2.9 Python

Python merupakan bahasa pemrograman yang populer dalam berbagai bidang, terutama dalam ilmu data dan pemodelan statistik. Kelebihan utama *python* terletak pada kemudahan penggunaannya, fleksibilitas, serta dukungan dari komunitas sumber terbuka yang luas. Bahasa ini memiliki berbagai pustaka yang memungkinkan analisis data yang efisien, termasuk dalam pemodelan deret waktu dan pembelajaran mesin.

Dalam konteks analisis deret waktu, *python* menyediakan berbagai alat yang mendukung pemrosesan data, peramalan, serta pengembangan model hibrida yang menggabungkan teknik statistik dan kecerdasan buatan. Model seperti ARIMA (*Autoregressive Integrated Moving Average*) sering digunakan untuk memprediksi pola jangka panjang dalam data historis, sedangkan *Long Short-Term Memory* (LSTM) digunakan untuk menangkap pola nonlinier dalam data volatil. Kombinasi kedua model ini, dikenal sebagai ARIMA-LSTM, telah menunjukkan hasil yang lebih akurat dibandingkan model statistik konvensional.

Untuk implementasi model semacam itu, *Python* menyediakan pustaka seperti *NumPy* untuk manipulasi *array*, *Pandas* untuk pemrosesan data, *Scikit-*

learn untuk pembelajaran mesin, serta *TensorFlow* dan Keras untuk pengembangan model berbasis jaringan saraf. Model ini juga dapat ditingkatkan dengan teknik lain seperti *Random forest*, yang membantu dalam pemilihan fitur dan optimalisasi parameter model.

Melalui pendekatan ini, *Python* tidak hanya menjadi alat utama dalam pemodelan deret waktu, tetapi juga memungkinkan integrasi dengan teknik pembelajaran mendalam lainnya, seperti *Convolutional Neural Network* (CNN) dan *Gated Recurrent Unit* (GRU), guna meningkatkan performa peramalan. Dengan ekosistem yang terus berkembang, *Python* tetap menjadi pilihan utama bagi ilmuwan data dalam mengembangkan model yang kompleks dan akurat (Lama et al., 2024).

