

**RANCANG BANGUN SISTEM KLASIFIKASI SENYAWA
DALAM TANAMAN YANG BERPOTENSIAL
MENYEMBUHKAN KANKER PARU-PARU
MENGUNAKAN ALGORITMA
RANDOM FOREST**



SKRIPSI

**KENNY MATTHEW
00000057330**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**RANCANG BANGUN SISTEM KLASIFIKASI SENYAWA
DALAM TANAMAN YANG BERPOTENSIAL
MENYEMBUHKAN KANKER PARU-PARU
MENGUNAKAN ALGORITMA
RANDOM FOREST**



SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

**KENNY MATTHEW
00000057330**

**UNIVERSITAS
MULTIMEDIA
NUSANTARA
PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Kenny Matthew

Nomor Induk Mahasiswa : 00000057330

Program Studi : Informatika

Skripsi dengan judul:

Rancang Bangun Sistem Klasifikasi Senyawa dalam Tanaman yang Berpotensi Menyembuhkan Kanker Paru-Paru Menggunakan Algoritma Random Forest

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 26 Juni 2025



(Kenny Matthew)

HALAMAN PENGESAHAN

Skripsi dengan judul
**RANCANG BANGUN SISTEM KLASIFIKASI SENYAWA
DALAM TANAMAN YANG BERPOTENSIAL
MENYEMBUHKAN KANKER PARU-PARU
MENGUNAKAN ALGORITMA
RANDOM FOREST**

oleh

Nama : Kenny Matthew
NIM : 00000057330
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Senin, 21 Juli 2025

Pukul 08.00 s/s 10.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang

Penguji

(David Agustriawan, S.Kom., M.Sc., Ph.D.) (Moeljono Widjaja, B.Sc., M.Sc., Ph.D.)
NIDN: 0525088601 NIDN: 0311106903

Pembimbing

(Angga Aditya Permana, S.Kom., M.Kom.)

NIDN: 0407128901

Ketua Program Studi Informatika,

(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)

NIDN: 0315109103

iii

Rancang Bangun Sistem..., Kenny Matthew, Universitas Multimedia Nusantara

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama : Kenny Matthew
NIM : 00000057330
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Rancang Bangun Sistem Klasifikasi
Senyawa dalam Tanaman yang
Berpotensial Menyembuhkan Kanker
Paru-Paru Menggunakan Algoritma
Random Forest

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan; dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - Embargo publikasi karya ilmiah dalam kurun waktu tiga tahun.

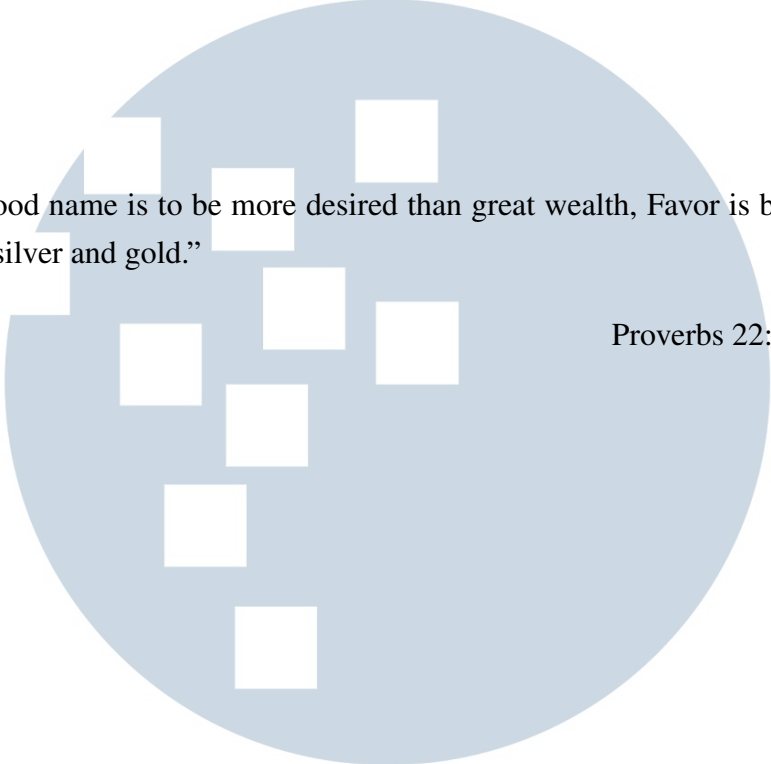
Tangerang, 26 Juni 2025

Yang menyatakan



Kenny Matthew

HALAMAN PERSEMBAHAN / MOTTO



"A good name is to be more desired than great wealth, Favor is better than silver and gold."

Proverbs 22:1 (NASB)

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa karena atas rahmat dan karunia-Nya, penulis dapat menyelesaikan penelitian ini dengan baik dan tepat waktu. Laporan ini disusun sebagai salah satu syarat dalam menyelesaikan tugas akademik di Teknik Informatika, Universitas Multimedia Nusantara. Penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Angga Aditya Permana, S.Kom., M.Kom., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga karya ilmiah ini dapat memberikan manfaat bagi pembaca serta menjadi kontribusi yang berarti dalam pengembangan ilmu pengetahuan, khususnya di bidang teknologi dan kesehatan.

Tangerang, 26 Juni 2025

UNIVERSITAS
MULTIMEDIA
NUSANTARA



Kenny Matthew

**RANCANG BANGUN SISTEM KLASIFIKASI SENYAWA
DALAM TANAMAN YANG BERPOTENSIAL
MENYEMBUHKAN KANKER PARU-PARU
MENGUNAKAN ALGORITMA
RANDOM FOREST**

Kenny Matthew

ABSTRAK

Kanker paru-paru merupakan salah satu penyebab utama kematian di dunia, sehingga diperlukan metode penemuan kandidat obat yang cepat dan efisien. Penelitian ini bertujuan untuk membangun sebuah sistem klasifikasi berbasis machine learning untuk memprediksi potensi antikanker senyawa kimia terhadap target protein Epidermal Growth Factor Receptor (EGFR), yang merupakan target penting dalam terapi kanker paru-paru. Data bioaktivitas yang terdiri dari 10074 senyawa unik dikumpulkan secara terprogram dari database ChEMBL. Setiap senyawa direpresentasikan menggunakan fitur molecular fingerprints (ECFP4 1024-bit) untuk menangkap informasi struktural secara detail. Model klasifikasi dibangun menggunakan algoritma Random Forest dengan parameter class weight=balanced untuk menangani dataset yang tidak seimbang. Evaluasi model menggunakan metode 5-Fold Cross-Validation menunjukkan performa yang sangat baik dengan rata-rata akurasi mencapai 93 %. Hasil ini menunjukkan bahwa pendekatan komputasi menggunakan molecular fingerprints dan algoritma Random Forest sangat efektif untuk melakukan skrining virtual dan mengidentifikasi senyawa yang berpotensi sebagai agen terapi kanker paru-paru.

Kata kunci: bioinformatika, kanker paru-paru, machine learning, molecular fingerprint, random forest.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

**DESIGN AND CONSTRUCTION OF A CLASSIFICATION SYSTEM FOR
COMPOUNDS IN PLANTS THAT HAVE THE POTENTIAL TO CURE LUNG
CANCER USING THE RANDOM FOREST ALGORITHM**

Kenny Matthew

ABSTRACT

Lung cancer is one of the leading causes of death worldwide, requiring a rapid and efficient drug candidate discovery method. This study aims to develop a machine learning-based classification system to predict the anticancer potential of chemical compounds against the Epidermal Growth Factor Receptor (EGFR) protein target, which is an important target in lung cancer therapy. Bioactivity data consisting of 10,074 unique compounds were collected programmatically from the ChEMBL database. Each compound was represented using molecular fingerprints (ECFP4 1024-bit) to capture detailed structural information. A classification model was built using the Random Forest algorithm with the class weight=balanced parameter to handle the imbalanced dataset. Model evaluation using the 5-Fold Cross-Validation method showed excellent performance with an average accuracy of 93%. These results indicate that the computational approach using molecular fingerprints and the Random Forest algorithm is very effective for conducting virtual screening and identifying compounds with potential as lung cancer therapeutic agents.

Keywords: *bioinformatics, lung cancer, machine learning, molecular fingerprint, random forest.*



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR KODE	xiii
DAFTAR RUMUS	xiv
DAFTAR LAMPIRAN	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Permasalahan	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB 2 LANDASAN TEORI	6
2.1 Tanaman Obat dan Senyawa Bioaktif	6
2.2 Kanker dan Mekanisme Penyembuhannya	8
2.2.1 Peran EGFR dalam Kanker Paru-Paru	9
2.3 Machine Learning dan Data Mining	10
2.4 Algoritma Random Forest	12
2.4.1 Mekanisme Pemisahan Node pada Pohon Keputusan	12
2.4.2 Cara Kerja Algoritma Random Forest	13
2.4.3 Kelebihan dari Algoritma Random Forest	14
2.4.4 Kekurangan Algoritma Random Forest	14
2.5 Evaluasi Model Klasifikasi	15
BAB 3 METODOLOGI PENELITIAN	18
3.1 Pengumpulan Data	18
3.2 Pra-Pemrosesan Data	19
3.3 Pemilihan Algoritma Random Forest dan Hyperparameter tuning	24
3.4 Evaluasi Model	25
3.5 Perancangan Sistem	26
BAB 4 HASIL DAN DISKUSI	29
4.1 Hasil Penelitian	29
4.1.1 Hasil Pengumpulan Data	29
4.1.2 Hasil Pra-Pemrosesan Data	29
4.1.3 Hasil Implementasi dan Evaluasi Model Random Forest	31
4.1.4 Hasil Implementasi Prototipe Sistem	36
4.2 Diskusi	38
BAB 5 SIMPULAN DAN SARAN	40
5.1 Simpulan	40
5.2 Saran	41



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 2.1	Tabel confusion matrix	16
Tabel 4.1	Hasil evaluasi kinerja model random forest	31
Tabel 4.2	Confusion matrix model random forest	32



DAFTAR GAMBAR

Gambar 2.1	Lipinkis rule	7
Gambar 2.2	Gambar molekul dalam senyawa tanaman	9
Gambar 2.3	Tahapan data mining	12
Gambar 2.4	Diagram random forest	15
Gambar 3.1	Metodologi penelitian	18
Gambar 3.2	Pembagian data latih dan data uji	23
Gambar 3.3	Cross validation	26
Gambar 3.4	Ilustrasi hasil prediksi senyawa yang berpotensi	27
Gambar 3.5	Ilustrasi hasil prediksi senyawa yang tidak berpotensi	28
Gambar 4.1	Distribusi fitur kimiawi senyawa	30
Gambar 4.2	Hasil balancing dataset	31
Gambar 4.3	Confusion matrix	33
Gambar 4.4	Kurva roc dan auc	34
Gambar 4.5	Kepentingan fitur	36
Gambar 4.6	Ilustrasi hasil jika senyawa berpotensi menyembuhkan penyakit kanker paru-paru	37
Gambar 4.7	Ilustrasi hasil jika senyawa tidak berpotensi menyembuhkan penyakit kanker paru-paru	37



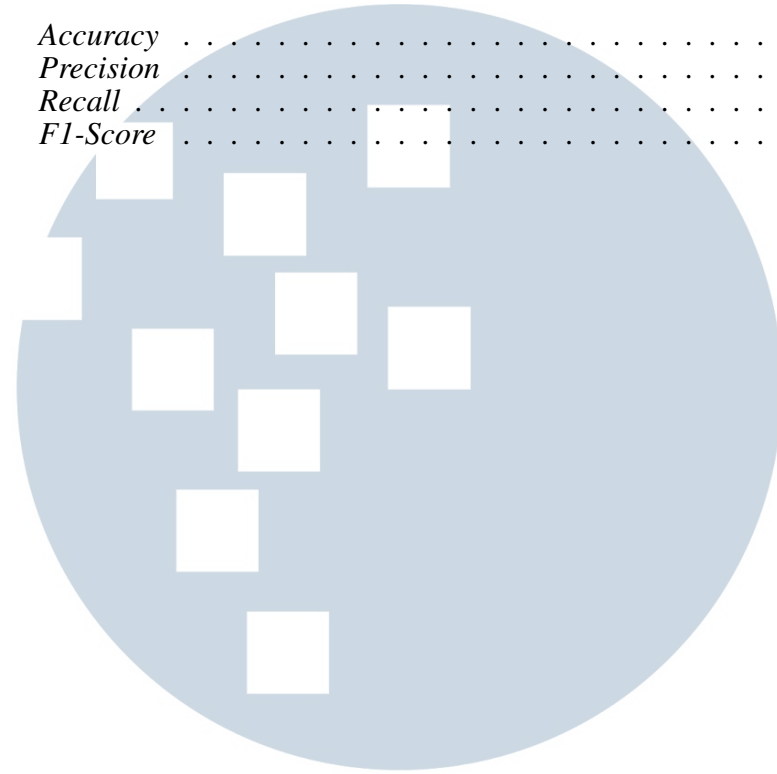
DAFTAR KODE

Kode 3.1	Load dataset	19
Kode 3.2	Menggabungkan fitur	20
Kode 3.3	Pembersihan data	21
Kode 3.4	Balancing dataset	21
Kode 3.5	Data latih dan data uji	23
Kode 3.6	Pemilihan model	24
Kode 3.7	Hyperparameter tuning	24
Kode 3.8	Model memprediksi	25
Kode 3.9	Laporan klasifikasi dan confusion matrix	25



DAFTAR RUMUS

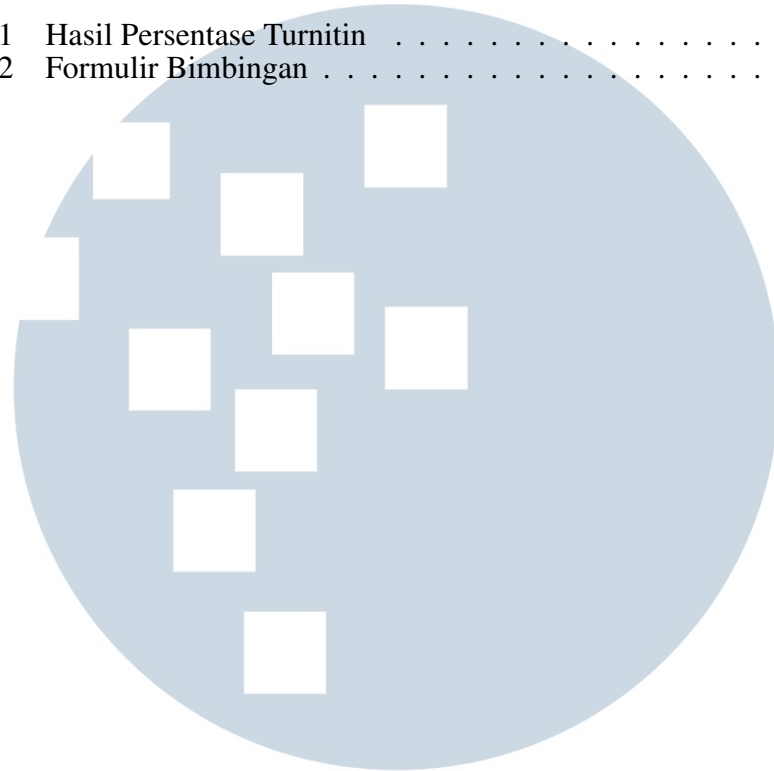
Rumus 2.1	<i>Accuracy</i>	16
Rumus 2.2	<i>Precision</i>	16
Rumus 2.3	<i>Recall</i>	17
Rumus 2.4	<i>F1-Score</i>	17



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran 1	Hasil Persentase Turnitin	45
Lampiran 2	Formulir Bimbingan	52



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA