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

SKRIPSI

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

KENNY MATTHEW
00000057330

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 Berpotensial 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 MENGGUNAKAN 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

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)

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

Kenny Matthew

RANCANG BANGUN SISTEM KLASIFIKASI SENYAWA DALAM TANAMAN YANG BERPOTENSIAL MENYEMBUHKAN KANKER PARU-PARU MENGGUNAKAN 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.

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

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR KODE

DAFTAR RUMUS

DAFTAR LAMPIRAN