KLASIFIKASI STADIUM KANKER PROSTAT MENGGUNAKAN LR DAN RF DENGAN SELEKSI FITUR LASSO-RFE BERBASIS DATA GEN DAN MIRNA

SKRIPSI

IVANDY WIJAYA
00000061844

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

KLASIFIKASI STADIUM KANKER PROSTAT MENGGUNAKAN LR DAN RF DENGAN SELEKSI FITUR LASSO-RFE BERBASIS DATA GEN DAN MIRNA

SKRIPSI

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

IVANDY WIJAYA
00000061844

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

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama	:	Ivandy Wijaya
Nomor Induk Mahasiswa	:	00000061844
Program Studi	:	Informatika

Skripsi dengan judul:

Klasifikasi Stadium Kanker Prostat Menggunakan LR-RF dengan Seleksi Fitur Lasso-RFE Berbasis Data Gen dan Mirna

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

(Ivandy Wijaya)

HALAMAN PENGESAHAN

Skripsi dengan judul

KLASIFIKASI STADIUM KANKER PROSTAT MENGGUNAKAN LR DAN RF DENGAN SELEKSI FITUR LASSO-RFE BERBASIS DATA GEN DAN MIRNA

oleh

Nama	:	Ivandy Wijaya
NIM	:	00000061844
Program Studi	:	Informatika
Fakultas	:	Fakultas Teknik dan Informatika

Telah diujikan pada hari Jumat, 18 Juli 2025
Pukul 15.00 s/d 17.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut

Ketua Sidang

(Dennis Gunawan, S.Kom., M.Sc.)
NIDN: 0320059001

Penguji

(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)
NIDN: 0315109103

Pembimbing I

(Moeljono Widjaja, B.Sc., M.Sc., Ph.D.)
NIDN: 0311106903

Pembimbing II

(David Agustriawan, S.Kom., M.Sc., Ph.D)
NIDN: 0525088601

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	:	Ivandy Wijaya
NIM	:	00000061844
Program Studi	:	Informatika
Jenjang	:	S1
Judul Karya Ilmiah	:	Klasifikasi Stadium Kanker Prostat Menggunakan LR dan RF dengan Seleksi Fitur Lasso-RFE Berbasis Data Gen dan Mirna

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

Ivandy Wijaya

**Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

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 Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini sebagai salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan laporan magang ini, sangatlah sulit bagi saya untuk menyelesaikan laporan magang ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Moeljono Widjaja, B.Sc., M.Sc., Ph.D., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Bapak David Agustriawan, S.Kom., M.Sc., Ph.D, sebagai Pembimbing kedua yang turut memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
6. 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 pengembangan ilmu pengetahuan, serta menjadi referensi bagi studi lanjutan di masa mendatang.

Tangerang, 26 Juni 2025

Ivandy Wijaya

KLASIFIKASI STADIUM KANKER PROSTAT MENGGUNAKAN LR DAN RF DENGAN SELEKSI FITUR LASSO-RFE BERBASIS DATA GEN DAN MIRNA

Ivandy Wijaya

ABSTRAK

Kanker prostat merupakan salah satu jenis kanker yang paling umum terjadi pada pria dan menghadirkan tantangan dalam proses penentuan stadium kanker yang dini dan akurat. Penelitian ini bertujuan untuk menggunakan pendekatan berbasis pembelajaran mesin dalam mengklasifikasikan kanker prostat stadium II dan III menggunakan data transkriptomik (RNA-seq dan miRNA) dari *The Cancer Genome Atlas Prostate Adenocarcinoma* (TCGA-PRAD). Dua pendekatan seleksi fitur digunakan, yaitu analisis gen yang diekspresikan secara berbeda (Differentially Expressed Genes/DEG) dan penyaringan statistik (LASSO + RFE). Delapan skenario eksperimen dikembangkan dengan menggabungkan berbagai jenis data, metode seleksi fitur, dan algoritma klasifikasi (Logistic Regression dan Random Forest). Performa terbaik diperoleh dari kombinasi data ekspresi gen dengan metode seleksi fitur LASSO + RFE serta algoritma Logistic Regression. Model yang dihasilkan mencapai akurasi sebesar 99,17%, precision sebesar 99,05%, recall sebesar 90%, dan F1-score sebesar 90%. Sebanyak 25 gen terpilih digunakan dalam model optimal akhir. Analisis lanjutan menunjukkan bahwa meskipun kemampuan diskriminatif tiap gen secara individu terbatas, penggunaan gabungan gen-gen tersebut memberikan performa klasifikasi yang unggul, yang menyoroti pentingnya interaksi antar gen.

Kata kunci: *Ekspresi Gen*, *Kanker prostat*, *miRNA*, *Pembelajaran mesin*, *Seleksi fitur*

PROSTATE CANCER STAGE CLASSIFICATION USING LR AND RF WITH LASSO-RFE FEATURE SELECTION BASED ON GENE AND MIRNA DATA

Ivandy Wijaya

ABSTRACT

Prostate cancer is one of the most prevalent cancers among men and presents significant challenges in early and accurate staging. This study proposes a machine learning-based approach to classify stage II and III prostate cancer using transcriptomic data (RNA-seq and miRNA) from the The Cancer Genome Atlas Prostate Adenocarcinoma (TCGA-PRAD) cohort. Two different feature selection approaches were employed, involving Differentially Expressed Genes (DEG) analysis and statistical filtering (LASSO + RFE). Eight experimental scenarios were developed by combining different data types, feature selection methods, and classification algorithms (Logistic Regression and Random Forest). The best performance was observed in the combination of gene expression data with the LASSO + RFE feature selection approach and the Logistic Regression algorithm. The resulting model achieved 99.17% accuracy, 99.05% precision, 90% recall, and a 90% F1-score. A set of 25 selected genes was used in the final optimal model. Further analysis revealed limited discriminative power individually, but collective use of selected genes provided superior classification performance, highlighting the importance of gene interactions.

Keywords: *Feature selection*, *Gene Expression*, *Machine learning*, *miRNA*, *Prostate cancer*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR KODE

DAFTAR RUMUS

DAFTAR LAMPIRAN