PREDIKSI STADIUM KANKER PROSTAT BERDASARKAN EKSPRESI GEN-MIRNA MENGGUNAKAN SVM DAN LR DENGAN SELEKSI FITUR DESEQ2-RFE

SKRIPSI

GREGORIUS IVAN HALIM
00000054295

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

PREDIKSI STADIUM KANKER PROSTAT BERDASARKAN EKSPRESI GEN-MIRNA MENGGUNAKAN SVM DAN LR DENGAN SELEKSI FITUR DESEQ2-RFE

SKRIPSI

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

GREGORIUS IVAN HALIM
00000054295

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

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama	:	Gregorius Ivan Halim
Nomor Induk Mahasiswa	:	00000054295
Program Studi	:	Informatika

Skripsi dengan judul:

Prediksi Stadium Kanker Prostat Berdasarkan Ekspresi Gen-miRNA menggunakan SVM dan LR dengan Seleksi Fitur DESeq2-RFE

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, 01 Juli 2025

(Gregorius Ivan Halim)

HALAMAN PENGESAHAN

Skripsi dengan judul

PREDIKSI STADIUM KANKER PROSTAT BERDASARKAN EKSPRESI GEN-MIRNA MENGGUNAKAN SVM DAN LR DENGAN SELEKSI FITUR DESEQ2-RFE

oleh

Nama	:	Gregorius Ivan Halim
NIM	:	00000054295
Program Studi	:	Informatika
Fakultas	:	Fakultas Teknik dan Informatika

Telah diujikan pada hari Jumat, 18 Juli 2025
Pukul 08.00 s/d 10.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut

Ketua Sidang	Penguji
(Moeljono Widjaja, B.Sc., M.Sc., Ph.D.)	(Dr. Maria Irmina Prasetiyowati, S.Kom., M.T.)
NIDN: 0311106903	NIDN: 0725057201
Pembimbing I	Pembimbing II
(Marlinda Vasty Overbeek, S.Kom, M.Kom)	(David Agustriawan, S.Kom., M.Sc.,Ph.D)
NIDN: 0818038501	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	:	Gregorius Ivan Halim
NIM	:	00000054295
Program Studi	:	Informatika
Jenjang	:	S1
Judul Karya Ilmiah	:	Prediksi Stadium Kanker Prostat Berdasarkan Ekspresi Gen-miRNA menggunakan SVM dan LR dengan Seleksi Fitur DESeq2-RFE

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, 01 Juli 2025

Yang menyatakan

Gregorius Ivan Halim

HALAMAN PERSEMBAHAN / MOTTO

”It’s not the hours you put in your work that counts, it’s the work you put in the hours.”

Sam Ewing

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat-Nya sehingga Tugas Akhir ini dapat diselesaikan. Tugas ini disusun sebagai syarat kelulusan Program Studi Informatika, Universitas Multimedia Nusantara. Penelitian ini mengembangkan model prediksi kanker prostat berbasis data gen dan miRNA dengan pendekatan komputasional. Terima kasih kepada semua pihak yang telah mendukung penyusunan tugas ini diantaranya:

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. Ibu Marlinda Vasty Overbeek, S.Kom, M.Kom, selaku Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi.
5. Bapak David Agustiawan, S.Kom., M.Sc.,Ph.D, selaku Pembimbing kedua, yang turut memberikan bimbingan dan masukan.
6. Keluarga saya yang selalu memberikan bantuan, dukungan moral maupun material.
7. Serta semua pihak yang tidak dapat disebutkan satu per satu namun telah membantu dalam proses penyusunan tugas akhir ini.

Penulis berharap karya ilmiah ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya dalam penelitian terkait deteksi kanker prostat, serta menjadi referensi bagi studi lanjutan di masa mendatang.

Tangerang, 01 Juli 2025

Gregorius Ivan Halim

PREDIKSI STADIUM KANKER PROSTAT BERDASARKAN EKSPRESI GEN-MIRNA MENGGUNAKAN SVM DAN LR DENGAN SELEKSI FITUR DESEQ2-RFE

Gregorius Ivan Halim

ABSTRAK

Klasifikasi stadium kanker prostat yang akurat sangat penting untuk meningkatkan hasil pengobatan dan mendukung pengambilan keputusan klinis. Penelitian ini bertujuan mengembangkan model prediktif untuk klasifikasi stadium kanker prostat berdasarkan data ekspresi gen dan miRNA, dengan memanfaatkan kombinasi teknik seleksi fitur dan algoritma pembelajaran mesin. Fokus utama penelitian ini adalah membandingkan kemampuan prediktif dari kedua jenis data dalam membedakan antara kanker prostat stadium II dan III pada pasien pria non-Hispanik kulit putih. Proses seleksi fitur dilakukan secara bertahap, dimulai dengan metode filter seperti DESeq2, Limma, ANOVA, dan MRMR, kemudian dilanjutkan dengan teknik wrapper Recursive Feature Elimination (RFE). Model dibangun menggunakan algoritma Support Vector Machine (SVM) dan Logistic Regression (LR) serta dilatih pada data RNA-Seq yang telah distandarisasi. Evaluasi performa dilakukan dengan menggunakan metrik akurasi, presisi, recall, dan skor F1 makro. Hasil menunjukkan bahwa model berbasis ekspresi gen secara konsisten mengungguli model berbasis miRNA pada semua metrik evaluasi. Model terbaik diperoleh dari kombinasi DESeq2–RFE–SVM, dengan akurasi 96,1% dan skor F1 makro 96%, hanya menggunakan 18 fitur gen. Meskipun sebagian besar gen memiliki skor AUC rendah dan korelasi lemah hingga sedang terhadap stadium kanker ketika dievaluasi secara individu, integrasi multigen memungkinkan klasifikasi stadium yang sangat efektif. Temuan ini menegaskan bahwa fitur ekspresi gen lebih informatif dibandingkan miRNA dalam klasifikasi stadium kanker prostat. Ke depan, pengembangan klasifikasi multikelas, penerapan metode deep learning, penggunaan RFE secara langsung, dan pengolahan fitur pada sistem berperforma tinggi disarankan untuk meningkatkan akurasi dan relevansi klinis model.

Kata kunci: Biomarker, Ekspresi gen, Kanker prostat, *Machine learning*, miRNA, Seleksi fitur

PREDICTION OF PROSTATE CANCER STAGE BASED ON GENE-MIRNA EXPRESSION USING SVM AND LR WITH DESEQ2-RFE FEATURE SELECTION

Gregorius Ivan Halim

ABSTRACT

Accurate staging classification of prostate cancer is crucial for improving treatment outcomes and supporting clinical decision-making. This study aims to develop predictive models for prostate cancer staging using gene and miRNA expression data, employing a combination of feature selection techniques and machine learning algorithms. The primary objective is to compare the predictive capabilities of both data types in distinguishing between stage II and stage III prostate cancer in non-Hispanic white male patients. Feature selection was performed in a multi-step process, beginning with filter methods such as DESeq2, Limma, ANOVA, and MRMR, followed by the wrapper-based Recursive Feature Elimination (RFE) technique. Predictive models were built using Support Vector Machine (SVM) and Logistic Regression (LR) algorithms, trained on standardized RNA-Seq data. Model performance was evaluated using accuracy, precision, recall, and macro F1-score metrics. The results demonstrated that gene expression-based models consistently outperformed miRNA-based models across all evaluation metrics. The best performance was achieved by the DESeq2–RFE–SVM model, which reached 96.1% accuracy and a 96% macro F1-score using only 18 gene features. Although most individual genes showed low AUC scores and weak to moderate correlations with cancer stage, their collective integration enabled highly effective stage classification. These findings confirm that gene expression features are more informative than miRNA in the context of prostate cancer stage classification. Future work should consider exploring multiclass classification, incorporating deep learning approaches, applying RFE directly, and executing feature selection on higher-performance systems to further enhance model accuracy and clinical relevance.

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

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR KODE

DAFTAR RUMUS

DAFTAR LAMPIRAN