KLASIFIKASI STADIUM KANKER PAYUDARA BERBASIS GEN-MIRNA DENGAN SVM DAN RIDGE REGRESSION SERTA SELEKSI FITUR ANOVA-RFE

SKRIPSI

AGUSTINUS
00000053639

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

KLASIFIKASI STADIUM KANKER PAYUDARA BERBASIS GEN-MIRNA DENGAN SVM DAN RIDGE REGRESSION SERTA SELEKSI FITUR ANOVA-RFE

SKRIPSI

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

AGUSTINUS
00000053639

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

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama	:	Agustinus
Nomor Induk Mahasiswa	:	00000053639
Program Studi	:	Informatika

Skripsi dengan judul:

Klasifikasi Stadium Kanker Payudara Berbasis Gen-miRNA dengan SVM dan Ridge Regression serta Seleksi Fitur ANOVA-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

(Agustinus)

HALAMAN PENGESAHAN

Skripsi dengan judul

KLASIFIKASI STADIUM KANKER PAYUDARA BERBASIS GEN-MIRNA DENGAN SVM DAN RIDGE REGRESSION SERTA SELEKSI FITUR ANOVA-RFE

oleh

Nama	:	Agustinus
NIM	:	00000053639
Program Studi	:	Informatika
Fakultas	:	Fakultas Teknik dan Informatika

Telah diujikan pada hari Senin, 14 Juli 2025
Pukul 08.00 s/d 10.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut

Ketua Sidang

(Dr. Ir. Winarno, M.Kom.)
NIDN: 0330106002

Penguji

(Angga Aditya Permana, S.Kom., M.Kom.)
NIDN: 0407128901

Pembimbing I

(Marlinda Vasty Overbeek, S.Kom., M.Kom.)
NIDN: 0818038501

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	:	Agustinus
NIM	:	00000053639
Program Studi	:	Informatika
Jenjang	:	S1
Judul Karya Ilmiah	:	Klasifikasi Stadium Kanker Payudara Berbasis Gen-miRNA dengan SVM dan Ridge Regression serta Seleksi Fitur ANOVA-RFE

Menyatakan dengan sesungguhnya bahwa saya bersedia:

☑ 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*) **.

Tangerang, 01 Juli 2025

Yang menyatakan

Agustinus

**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

”Turn your fear into adrenaline to fuel your passion, utilize that time to get it done and prove that you can”

Agustinus

KATA PENGANTAR

Puji syukur ke hadirat Tuhan Yang Maha Esa atas terselesaikannya penulisan skripsi ini sebagai bagian dari pemenuhan persyaratan akademik pada Program Studi Informatika, Universitas Multimedia Nusantara. Penyusunan skripsi ini melalui berbagai proses yang penuh tantangan, refleksi, dan pembelajaran yang berharga. Ucapan terima kasih disampaikan kepada pihak-pihak yang telah memberikan dukungan, bantuan, dan bimbingan, antara lain:

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 dengan sabar membimbing dan meluangkan waktu serta pemikiran dalam setiap tahap penyusunan skripsi ini.
5. Bapak David Agustriawan, S.Kom., M.Sc., Ph.D., selaku Pembimbing Kedua yang telah memberikan masukan dan dukungan dalam penyusunan skripsi ini.
6. Keluarga yang senantiasa memberikan semangat, kepercayaan, serta dukungan moral dan material.

Skripsi ini diharapkan dapat berkontribusi nyata dalam pengembangan ilmu pengetahuan dan mendorong riset lanjutan di masa depan.

Tangerang, 01 Juli 2025

Agustinus

KLASIFIKASI STADIUM KANKER PAYUDARA BERBASIS GEN-MIRNA DENGAN SVM DAN RIDGE REGRESSION SERTA SELEKSI FITUR ANOVA-RFE

Agustinus

ABSTRAK

Kanker payudara merupakan penyebab utama kematian pada wanita, di mana deteksi stadium dini krusial untuk meningkatkan prognosis pasien. Penelitian ini bertujuan mengembangkan model klasifikasi berbasis *machine learning* untuk membedakan stadium I dan III kanker payudara menggunakan data ekspresi gen dari *The Cancer Genome Atlas* (TCGA) pada wanita ras kulit putih. Pendekatan yang digunakan melibatkan seleksi fitur dengan *Analysis of Variance* (ANOVA) untuk memilih 1000 fitur awal, dilanjutkan *Recursive Feature Elimination* (RFE) hingga diperoleh 35 fitur optimal, dan klasifikasi menggunakan *Logistic Regression* dengan penalti L2. Hasil penelitian menunjukkan model terbaik mampu mencapai akurasi 97%, presisi 97%, *recall* 97%, dan F1-score 97% pada skenario stadium I vs. III, dengan durasi komputasi hanya 8 detik. Performa ini lebih unggul dibandingkan skenario stadium II vs. III (F1-score maksimum 88%) dan penggunaan data miRNA (F1-score maksimum 82%), menegaskan efektivitas data ekspresi gen dalam klasifikasi stadium. Sebanyak 35 gen kandidat biomarker berhasil diidentifikasi, dengan 10 gen didukung literatur terkait kanker payudara, sementara 25 lainnya berpotensi sebagai penanda baru. Analisis ROC AUC individual menunjukkan nilai maksimum 0,639, mengindikasikan kekuatan prediktif bergantung pada kombinasi multivariat. Penelitian ini membuktikan potensi pendekatan pembelajaran mesin dalam mendeteksi stadium kanker payudara secara akurat dan efisien.

Kata Kunci: Ekspresi Gen, Kanker Payudara, Klasifikasi Stadium, *Machine learning*, Seleksi Fitur

BREAST CANCER STAGE CLASSIFICATION BASED ON MIRNA GENES USING SVM AND RIDGE REGRESSION WITH ANOVA-RFE FEATURE SELECTION

Agustinus

ABSTRACT

Breast cancer is a leading cause of death in women, where early stage detection is crucial to improve patient prognosis. This study aims to develop a machine learning-based classification model to distinguish between stages I and III of breast cancer using gene expression data from The Cancer Genome Atlas (TCGA) in white women. The approach used involved feature selection using Analysis of Variance (ANOVA) to select 1000 initial features, followed by Recursive Feature Elimination (RFE) to obtain 35 optimal features, and classification using Logistic Regression with L2 penalty. The results showed that the best model was able to achieve 97% accuracy, 97% precision, 97% recall, and F1-score. 97%, and F1-score 97% in the stage I vs. III scenario, with a computation duration of only 8 seconds. This performance is superior to the stage II vs. III scenario (maximum F1-score 88%) and the use of miRNA data (maximum F1-score 82%), confirming the effectiveness of gene expression data in stage classification. A total of 35 biomarker candidate genes were identified, with 10 supported by breast cancer-related literature, while the other 25 have potential as novel markers. ROC analysis of individual AUCs showed a maximum value of 0.639, indicating predictive power was dependent on multivariate combinations. This research proves the potential of machine learning approaches in detecting breast cancer stages accurately and efficiently.

Keywords: *Gene Expression, Breast Cancer, Stage Classification, Machine learning, Feature Selection*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR KODE

DAFTAR RUMUS

DAFTAR LAMPIRAN