PREDIKSI STADIUM KANKER PAYUDARA BERBASIS DATA MIRNA DAN GEN DENGAN SELEKSI FITUR DAN VOTING CLASSIFIER

SKRIPSI

RENFRED LEEMAN
00000056836

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

PREDIKSI STADIUM KANKER PAYUDARA BERBASIS DATA MIRNA DAN GEN DENGAN SELEKSI FITUR DAN VOTING CLASSIFIER

SKRIPSI

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

RENFRED LEEMAN
00000056836

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

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama	:	Renfred Leeman
Nomor Induk Mahasiswa	:	00000056836
Program Studi	:	Informatika

Skripsi dengan judul:

Prediksi Stadium Kanker Payudara Berbasis Data miRNA dan Gen dengan Seleksi Fitur dan Voting Classifier

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

(Renfred Leeman)

HALAMAN PENGESAHAN

Skripsi dengan judul

PREDIKSI STADIUM KANKER PAYUDARA BERBASIS DATA MIRNA DAN GEN DENGAN SELEKSI FITUR DAN VOTING CLASSIFIER

oleh

Nama	:	Renfred Leeman
NIM	:	00000056836
Program Studi	:	Informatika
Fakultas	:	Fakultas Teknik dan Informatika

Telah diujikan pada hari Jumat, 18 Juli 2025
Pukul 10.00 s/d 12.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut

Ketua Sidang	Penguji
(Dr. Maria Irmina Prasetiyowati, S.Kom., M.T.)	(Moeljono Widjaja, B.Sc., M.Sc., Ph.D)
NIDN: 0725057201	NIDN: 0311106903

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	:	Renfred Leeman
NIM	:	00000056836
Program Studi	:	Informatika
Jenjang	:	S1
Judul Karya Ilmiah	:	Prediksi Stadium Kanker Payudara Berbasis Data miRNA dan Gen dengan Seleksi Fitur dan Voting Classifier

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

Renfred Leeman

**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

”Überm Sternenzelt richtet Gott, wie wir gerichtet.”

Friedrich Schiller

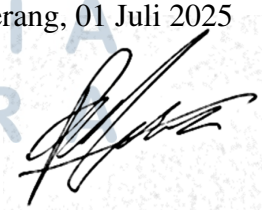
KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga skripsi yang berjudul ”Prediksi Stadium Kanker Payudara Berbasis Data miRNA dan Gen dengan Seleksi Fitur dan Voting Classifier” ini dapat diselesaikan dengan baik. Penulis 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. Ibu Marlinda Vasty Overbeek, S.Kom, M.Kom, 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 telah 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.

Harapannya, penelitian ini dapat memberikan kontribusi dalam pengembangan metode deteksi dini kanker payudara serta menjadi acuan bagi penelitian lanjutan di bidang bioinformatika dan ilmu kedokteran.

Tangerang, 01 Juli 2025



Renfred Leeman

PREDIKSI STADIUM KANKER PAYUDARA BERBASIS DATA MIRNA DAN GEN DENGAN SELEKSI FITUR DAN VOTING CLASSIFIER

Renfred Leeman

ABSTRAK

Kanker payudara merupakan salah satu penyebab utama kematian pada perempuan di seluruh dunia. Deteksi dini terhadap stadium kanker payudara sangat penting untuk meningkatkan peluang keberhasilan pengobatan. Penelitian ini bertujuan untuk membangun model klasifikasi stadium kanker payudara (*early stage* dan *advanced stage*) dengan memanfaatkan data ekspresi gen dan miRNA dari basis data *The Cancer Genome Atlas* (TCGA). Proses penelitian melibatkan tahapan praproses data, seleksi gen dan miRNA berdasarkan dan tanpa analisis *Differentially Expressed Genes* (DEG) menggunakan metode *limma*, dilanjutkan dengan seleksi fitur menggunakan kombinasi *Logistic Regression* (L1) dan *Recursive Feature Elimination* (RFE). Model klasifikasi dilatih menggunakan algoritma *Support Vector Machine* (SVM), *Random Forest* (RF), *Logistic Regression* (LR), dan *Voting Classifier*. Evaluasi dilakukan secara realistis menggunakan pembagian data *train* dan *test*. Hasil penelitian menunjukkan bahwa data ekspresi gen dengan model *Voting Classifier* mampu mencapai akurasi hingga 92,6% pada data pengujian. Selain itu, penelitian ini berhasil mengidentifikasi kombinasi biomarker potensial sebanyak 41 fitur, yang menunjukkan kontribusi kolektif yang lebih kuat dibandingkan penggunaan fitur tunggal. Temuan ini menegaskan bahwa seleksi fitur terstruktur dan pendekatan *multivariabel* dapat meningkatkan akurasi klasifikasi stadium kanker payudara serta mendukung identifikasi biomarker molekuler secara lebih efisien dan aplikatif.

Kata kunci: Biomarker, Differentially Expressed Genes, Gen, Kanker payudara, miRNA

BREAST CANCER STAGE PREDICTION BASED ON MIRNA AND GENE EXPRESSION DATA USING FEATURE SELECTION AND VOTING CLASSIFIER

Renfred Leeman

ABSTRACT

Breast cancer is one of the leading causes of mortality among women worldwide. Early detection of breast cancer stages is crucial to improve treatment success rates. This study aims to develop a classification model to distinguish breast cancer stages (early stage and advanced stage) using gene expression and miRNA data from The Cancer Genome Atlas (TCGA) database. The research process involved data preprocessing, selection of genes and miRNAs with and without Differentially Expressed Genes (DEG) analysis using the limma method, followed by feature selection combining Logistic Regression (L1) and Recursive Feature Elimination (RFE). Classification models were trained using Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), and Voting Classifier algorithms. Evaluation was conducted realistically using train-test data splits. The results demonstrated that gene expression data classified with the Voting Classifier achieved an accuracy of up to 92.6% on the test set. Furthermore, this study identified a potential biomarker combination consisting of 41 features, which collectively provided stronger predictive power compared to single-feature approaches. These findings highlight that structured feature selection and multivariable approaches can improve the accuracy of breast cancer stage classification while supporting a more efficient and practical identification of molecular biomarkers.

Keywords: *Biomarkers, Breast cancer, Differentially Expressed Genes, Gene, miRNA*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR KODE

DAFTAR RUMUS

DAFTAR LAMPIRAN