

**KLASIFIKASI STAGING KANKER GASTRIC
MENGUNAKAN MACHINE LEARNING PADA
DATASET GENOMIC GENE EXPRESSION**



LAPORAN RESEARCH & TECHNOLOGY PROGRAM

**CAESAR EKA NATHANAEL
00000067638**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2026**

**KLASIFIKASI STAGING KANKER GASTRIC
MENGUNAKAN MACHINE LEARNING PADA
DATASET GENOMIC GENE EXPRESSION**



LAPORAN RESEARCH & TECHNOLOGY PROGRAM

**CAESAR EKA NATHANAEL
00000067638**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2026**

HALAMAN PERNYATAAN ORISINALITAS TIDAK PLAGIAT

Dengan ini saya,

Nama : Caesar Eka Nathanael

NIM : 00000067638

Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa Laporan Research & Technology Program saya yang berjudul:

Klasifikasi Staging Kanker Gastric Menggunakan Machine Learning pada Dataset Genomic Gene Expression

merupakan hasil karya saya sendiri, bukan merupakan hasil plagiat, dan tidak pula dituliskan oleh orang lain; Semua sumber, baik yang dikutip maupun dirujuk, telah saya cantumkan dan nyatakan dengan benar pada bagian Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan karya ilmiah, saya bersedia menerima konsekuensi untuk dinyatakan TIDAK LULUS. Saya juga bersedia menanggung segala konsekuensi hukum yang berkaitan dengan tindak plagiarisme ini sebagai kesalahan saya pribadi dan bukan tanggung jawab Universitas Multimedia Nusantara.

Tangerang, 02 Januari 2026

UNIVERSITAS
MULTIMEDIA
NUSANTARA



(Caesar Eka Nathanael)

HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)

Saya yang bertanda tangan di bawah ini:

Nama : Caesar Eka Nathanael
NIM : 00000067638
Program Studi : Informatika
Judul Laporan : Klasifikasi Staging Kanker Gastric Menggunakan Machine Learning pada Dataset Genomic Gene Expression

Dengan ini saya menyatakan secara jujur menggunakan bantuan Kecerdasan Artifisial (AI) dalam pengerjaan Laporan sebagai berikut :

- ☐ Menggunakan AI sebagaimana diizinkan untuk membantu dalam menghasilkan ide-ide utama saja
- ☐ Menggunakan AI sebagaimana diizinkan untuk membantu menghasilkan teks pertama saja
- ☒ Menggunakan AI untuk menyempurnakan sintaksis dan tata bahasa untuk pengumpulan tugas
- ☐ Karena tidak diizinkan: Tidak menggunakan bantuan AI dengan cara apa pun dalam pembuatan tugas

Saya juga menyatakan bahwa:

- (1) Menyerahkan secara lengkap dan jujur penggunaan perangkat AI yang diperlukan dalam tugas melalui Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI)
- (2) Mengakui telah menggunakan bantuan AI dalam tugas saya baik dalam bentuk kata, paraphrase, penyertaan ide atau fakta penting yang disarankan oleh AI dan saya telah menyantumkan dalam sitasi serta referensi
- (3) Terlepas dari pernyataan di atas, tugas ini sepenuhnya merupakan karya saya sendiri

Tangerang, 02 Januari 2026


Caesar Eka Nathanael

HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN

Nama : Caesar Eka Nathanael
NIM : 00000067638
Program Studi : Informatika
Fakultas : Teknik dan Informatika

menyatakan bahwa saya melaksanakan kegiatan di:

Nama Perusahaan/Organisasi : LPPM Universitas Multimedia Nusantara
Alamat : Jl. Scientia Boulevard Gading, Curug
Sangereng, Serpong, Kabupaten Tangerang,
Banten 15810
Email Perusahaan/Organisasi : lppm@umn.ac.id

1. Perusahaan/Organisasi tempat saya melakukan kegiatan dapat divalidasi keberadaannya.
2. Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan data yang tidak valid di perusahaan/organisasi tempat saya melakukan kegiatan, maka:
 - a. Saya bersedia menerima konsekuensi dinyatakan **TIDAK LULUS**.
 - b. Saya bersedia menerima semua sanksi sesuai peraturan Universitas Multimedia Nusantara.

Pernyataan ini saya buat dengan sebenar-benarnya dan digunakan sebagaimana mestinya.

Tangerang, 19 November 2025

Mengetahui



DAVID AGUSTRIAWAN, S.KOM.,
M.SC., PH.D.



Menyatakan



Caesar Eka Nathanael

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH MAHASISWA

Yang bertanda tangan di bawah ini:

Nama : Caesar Eka Nathanael
NIM : 00000067638
Program Studi : Informatika
Jenjang : S1
Jenis Karya : Laporan Research & Technology Program

Menyatakan dengan sesungguhnya bahwa:

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk memublikasikan hasil laporan saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh civitas academica UMN/publik. Saya menyatakan bahwa laporan yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia memublikasikan hasil laporan ini ke dalam repositori Knowledge Center karena ada data yang bersifat konfidensial.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara.
 - ☐ Embargo publikasi laporan dalam kurun waktu 3 tahun.

Tangerang, 02 Januari 2026



Caesar Eka Nathanael

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

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

Halaman Persembahan / Motto

"Let us not lose heart in doing good, for in due time we will reap if we do not grow weary."

Galatians 6:9



KATA PENGANTAR

Puji syukur dipanjatkan kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga laporan MBKM Penelitian berjudul **KLASIFIKASI STAGING KANKER GASTRIC MENGGUNAKAN MACHINE LEARNING PADA DATASET GENOMIC GENE EXPRESSION** ini dapat diselesaikan dengan baik. Laporan ini disusun sebagai salah satu syarat untuk mencapai gelar Sarjana Ilmu Komputer pada Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara. Ucapan terima kasih disampaikan kepada seluruh pihak yang telah memberikan dukungan dan bimbingan selama pelaksanaan penelitian serta penyusunan laporan ini.

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 Vincentius Kurniawan, S.KOM., M.ENG.SC., selaku dosen pembimbing yang telah memberikan arahan dan bimbingan selama penelitian.
5. Orang tua dan keluarga atas dukungan serta doa yang telah diberikan.

Disadari bahwa laporan ini masih memiliki keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan. Semoga laporan ini dapat memberikan manfaat bagi pembaca serta mendukung pengembangan penelitian selanjutnya.

Tangerang, 02 Januari 2026



Caesar Eka Nathanael

KLASIFIKASI STAGING KANKER GASTRIC MENGGUNAKAN MACHINE LEARNING PADA DATASET GENOMIC GENE EXPRESSION

Caesar Eka Nathanael

ABSTRAK

Kanker lambung merupakan salah satu penyebab utama kematian akibat kanker di dunia, yang tingginya angka mortalitas sebagian besar dipengaruhi oleh keterlambatan diagnosis serta tantangan dalam penentuan stadium penyakit secara akurat. Penentuan stadium yang tepat sangat berperan dalam pemilihan terapi dan penentuan prognosis pasien. Penelitian ini bertujuan untuk mengembangkan model klasifikasi stadium kanker lambung berbasis data ekspresi gen menggunakan pendekatan *machine learning*. Data yang digunakan berupa data ekspresi gen *RNA-seq* berdimensi tinggi yang telah melalui tahap pra-pemrosesan dan pemeriksaan kualitas. Seleksi fitur dilakukan menggunakan metode regularisasi *Elastic Net* untuk mengidentifikasi gen-gen yang paling relevan terhadap status stadium kanker. Beberapa ukuran subset fitur dievaluasi guna menganalisis pengaruh jumlah fitur terhadap kinerja model. Selanjutnya, berbagai algoritma pembelajaran mesin diterapkan, meliputi *Logistic Regression*, *Support Vector Machine (SVM)*, *Random Forest*, *Multilayer Perceptron (MLP)*, serta *Voting Ensemble*. Optimasi hiperparameter dilakukan menggunakan metode *Random Search* dengan skema validasi silang. Evaluasi kinerja model dilakukan menggunakan metrik *ROC-AUC* dan analisis *confusion matrix*. Hasil penelitian menunjukkan bahwa model *SVM* dan *Voting Ensemble* memberikan performa yang kompetitif dan stabil dalam membedakan stadium awal dan lanjut kanker lambung. Hasil ini menunjukkan potensi pendekatan *machine learning* berbasis ekspresi gen sebagai dasar pengembangan sistem pendukung keputusan klinis.

Kata kunci: Ekspresi gen, Kanker lambung, *Machine learning*, Seleksi fitur, *Staging*

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

GASTRIC CANCER CLASSIFICATION USING MACHINE LEARNING ON GENOMIC GENE EXPRESSION DATASETS

Caesar Eka Nathanael

ABSTRACT

Gastric cancer is one of the leading causes of cancer-related mortality worldwide, largely due to delayed diagnosis and challenges in accurately determining disease stage. Precise cancer staging plays a crucial role in therapy selection and prognosis assessment. Therefore, this study aims to develop a machine learning-based gastric cancer stage classification model using gene expression data. The dataset consists of high-dimensional RNA-seq gene expression data that have undergone preprocessing and quality control procedures. Feature selection was performed using Elastic Net regularization to identify a subset of genes most relevant to cancer stage status. Multiple feature subset sizes were evaluated to investigate the impact of the number of selected genes on model performance. Subsequently, several machine learning algorithms were applied, including Logistic Regression, Support Vector Machine (SVM), Random Forest, Multilayer Perceptron (MLP), and a Voting Ensemble. Hyperparameter optimization was conducted using Random Search combined with cross-validation. Model performance was evaluated using the Receiver Operating Characteristic–Area Under the Curve (ROC-AUC) metric and confusion matrix analysis. The results demonstrate that the SVM and Voting Ensemble models achieved competitive and stable classification performance in distinguishing early-stage and advanced-stage gastric cancer. These findings suggest that machine learning approaches based on gene expression data, combined with appropriate feature selection, have the potential to serve as a foundation for developing clinical decision support systems to improve the accuracy of gastric cancer staging.

Keywords: Ensemble learning, Gastric cancer, Gene expression, Machine learning, Staging

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN ORISINALITAS	ii
HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)	iii
HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN	iv
HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN	iv
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
HALAMAN PERSEMBAHAN/MOTO	vi
KATA PENGANTAR	vii
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR RUMUS	xiv
DAFTAR LAMPIRAN	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Penelitian	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB 2 TINJAUAN PUSTAKA	5
2.1 Kanker Lambung (Gastric Cancer)	5
2.1.1 Stadium Kanker Lambung	5
2.1.2 Keterbatasan Metode Konvensional	6
2.2 Analisis Ekspresi Gen	6
2.3 Seleksi Fitur pada Data Ekspresi Gen	7
2.4 Seleksi Fitur Menggunakan LASSO	7
2.5 Machine Learning untuk Klasifikasi Stadium Kanker Lambung	8
2.6 Support Vector Machine	8
2.7 Multi-Layer Perceptron	9
2.8 Ensemble Learning	9
2.9 Metrik Evaluasi Model	9
BAB 3 METODE PENELITIAN	11
3.1 Alur Metodologi Penelitian	11
3.2 Metode Penelitian	12
3.3 Pendekatan Metodologis	13
3.4 Sumber dan Karakteristik Data	13
3.5 Pra-pemrosesan Data	14
3.5.1 Penyelarasan Data	14
3.5.2 Penanganan Nilai Hilang	14
3.5.3 Normalisasi dan Standarisasi	15
3.5.4 Transformasi Label	15
3.6 Seleksi Fitur	16
3.6.1 Seleksi Fitur Menggunakan LASSO	16
3.7 Pembagian Data	16

3.8	Pembangunan Model Klasifikasi	17
3.9	Optimasi Hiperparameter	17
3.10	Evaluasi Model	17
3.11	Analisis Feature Importance	18
3.12	Analisis Pengaruh Seleksi Fitur	18
3.13	Pipeline Eksperimen	18
3.14	Ringkasan Metodologi	18
BAB 4	HASIL DAN PEMBAHASAN	19
4.1	Hasil Perancangan Sistem	19
4.1.1	Hasil Pra-pemrosesan Data	19
4.1.2	Hasil Seleksi Fitur Menggunakan LASSO	20
4.1.3	Hasil Pembangunan Model Klasifikasi	21
4.1.4	Hasil Evaluasi Model Berdasarkan Variasi Jumlah Fitur	21
4.1.5	Confusion Matrix Model Terbaik	24
4.2	Pembahasan Hasil	25
4.2.1	Analisis Pengaruh Jumlah Fitur terhadap Performa Model	25
4.2.2	Analisis Karakteristik Model Klasifikasi	26
4.2.3	Analisis dari Perspektif Klinis	27
4.2.4	Perbandingan dengan Penelitian Terdahulu	27
4.2.5	Keterbatasan Penelitian	28
BAB 5	SIMPULAN SARAN	29
5.1	Simpulan	29
5.2	Saran	30
DAFTAR PUSTAKA		31



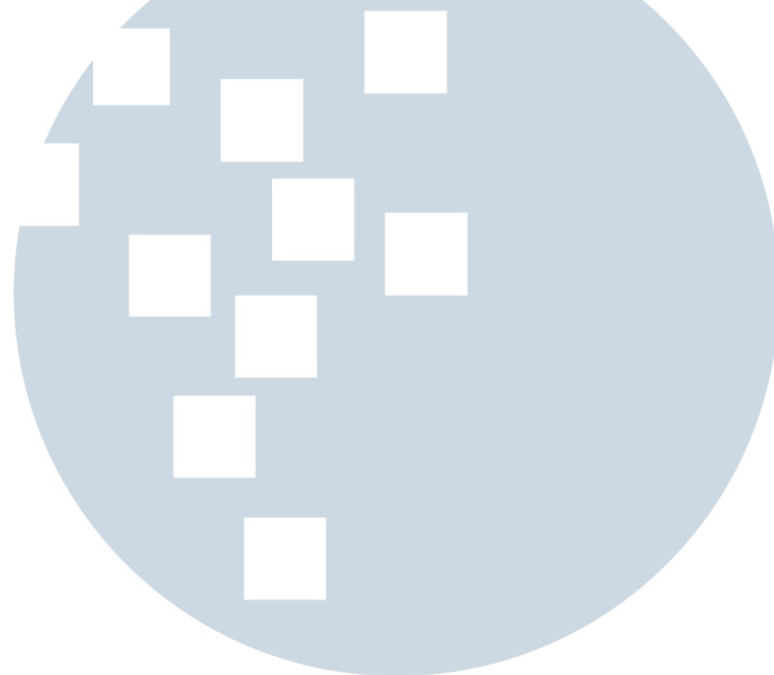
DAFTAR TABEL

Tabel 4.1	Hasil Evaluasi Model pada Data Uji Berdasarkan Variasi Jumlah Fitur	22
Tabel 4.2	Hasil Cross-Validation Model Berdasarkan Variasi Jumlah Fitur	23



DAFTAR GAMBAR

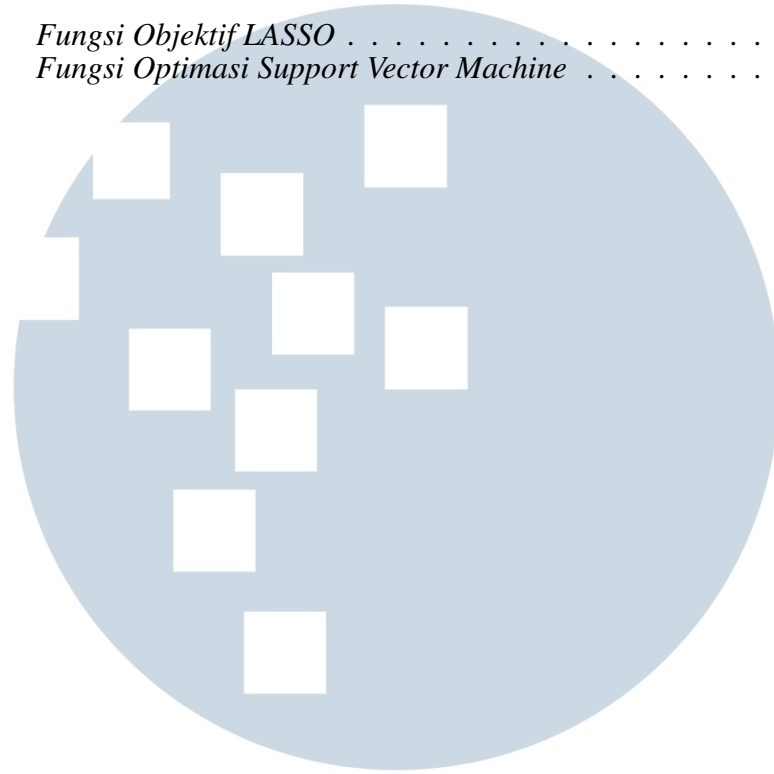
Gambar 3.1	Alur Metodologi Penelitian	12
Gambar 4.1	Confusion Matrix Model SVM dengan 50 Fitur	24



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR RUMUS

Rumus 2.1	<i>Fungsi Objektif LASSO</i>	8
Rumus 2.2	<i>Fungsi Optimasi Support Vector Machine</i>	8



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran 1	MBKM-01 Cover Letter MBKM Research	33
Lampiran 2	MBKM-02 MBKM Research Card	34
Lampiran 3	MBKM-03 Daily Task - MBKM Research	35
Lampiran 4	MBKM-04 Verification Form of Research Report MBKM Research	46
Lampiran 5	Letter of Acceptance	47
Lampiran 6	Form Bimbingan	48
Lampiran 7	Laporan Turnitin	49
Lampiran 8	Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI) . .	57
Lampiran 9	Surat Kerjasama Penelitian	58
Lampiran 10	Draft Research Paper	59

