

**SELEKSI FITUR EKSPRESI GEN MENGGUNAKAN  
FSACE UNTUK PREDIKSI KANKER USUS BESAR**



**LAPORAN RESEARCH TECHNOLOGY PROGRAM**

**STEVEN LIE  
00000069236**

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

**SELEKSI FITUR EKSPRESI GEN MENGGUNAKAN  
FSACE UNTUK PREDIKSI KANKER USUS BESAR**



**STEVEN LIE**

**00000069236**

**UMN**

**UNIVERSITAS**

**MULTIMEDIA**

**NUSANTARA**

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

**TANGERANG**

**2026**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Steven Lie  
Nomor Induk Mahasiswa : 00000069236  
Program Studi : Informatika

Skripsi dengan judul:

**Seleksi Fitur Ekspresi Gen Menggunakan FSACE Untuk Prediksi Kanker Usus Besar**

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 Januari 2026



Steven Lie

UMM  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)

Saya yang bertanda tangan dibawah ini,

Nama : Steven Lie  
NIM : 00000069236  
Program Studi : Informatika  
Judul Laporan : Seleksi Fitur Ekspresi Gen  
Menggunakan FSACE Untuk Prediksi  
Kanker Usus Besar

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

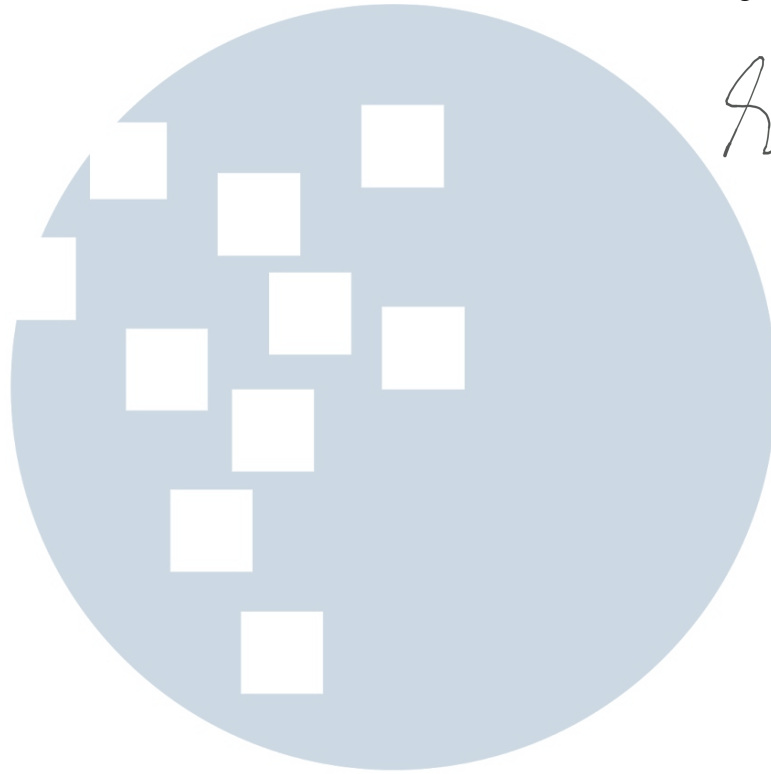
- ☒ Menggunakan AI sebagaimana diizinkan untuk membantu dalam menghasilkan ide-ide utama saja
- ☒ Menggunakan AI sebagaimana diizinkan untuk membantu menghasilkan teks pertama saja
- ☒ Menggunakan AI untuk menyempurnakan sintaksis dan tata bahasa untuk pengumpulan tugas
- ☐ Karena tidak diizinkan: Tidak 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, 01 Januari 2026

Yang menyatakan



A handwritten signature in black ink, which appears to read 'Steven Lie'.

Steven Lie

UMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN

Yang bertanda tangan dibawah ini:

Nama : Steven Lie  
NIM : 00000069236  
Program Studi : Informatika  
Fakultas : TEKNIK DAN INFORMATIKA

menyatakan bahwa saya melaksanakan kegiatan di:

Nama Perusahaan/Organisasi : Universitas Brawijaya  
Alamat : Jalan Veteran, Ketawanggede, Kecamatan  
Lowokwaru, Kota Malang, Jawa Timur  
Email Perusahaan/Organisasi : akademik.center@ub.ac.id

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

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

Mengetahui,



dr. Edvin Perwira Negara

Tangerang, 01 Januari 2026

Menyatakan



Steven Lie

## HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama : Steven Lie  
NIM : 00000069236  
Program Studi : Informatika  
Jenjang : S1  
Judul Karya Ilmiah : Seleksi Fitur Ekspresi Gen Menggunakan  
FSACE Untuk Prediksi Kanker Usus Besar

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 Januari 2026

Yang menyatakan



Steven Lie

UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

\*\*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**

"Trough chances various, trough all vicissidutes, we make our way."

Aeneid





## KATA PENGANTAR

Puji syukur panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga penelitian ini dapat diselesaikan dengan baik. Penelitian ini bertujuan untuk mengembangkan metode seleksi fitur ekspresi gen menggunakan FSACE untuk prediksi kanker usus besar. Dalam proses penyusunan penelitian ini, sebagai bentuk apresiasi atas segala dukungan yang diberikan, berikut disampaikan ucapan terima kasih kepada pihak-pihak terkait.

Mengucapkan terima kasih

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 Aditiyawan, S.Kom., M.Si., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Bapak Moeljono Widjaja, B.Sc., M.Sc., Ph.D., sebagai Pembimbing lapangan yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
6. Keluarga yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan laporan MBKM ini.
7. Teman-teman yang telah memberikan dukungan moral dalam menyelesaikan laporan MBKM ini.

Semoga karya ilmiah ini dapat memberikan manfaat bagi pembaca dan pengembangan ilmu pengetahuan.

Tangerang, 01 Januari 2026



Steven Lie

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

# SELEKSI FITUR EKSPRESI GEN MENGGUNAKAN FSACE UNTUK PREDIKSI KANKER USUS BESAR

Steven Lie

## ABSTRAK

Data ekspresi gen pada kanker kolon memiliki karakteristik dimensi yang sangat tinggi dengan jumlah sampel yang relatif terbatas, sehingga dapat menurunkan kinerja klasifikasi dan meningkatkan risiko *overfitting*. Oleh karena itu, seleksi fitur diperlukan untuk mengidentifikasi gen yang relevan. Penelitian ini mengusulkan penggunaan *Feature Selection using Approximate Conditional Entropy* (FSACE) untuk mereduksi dimensi data ekspresi gen kanker kolon. Data *RNA-Seq* diperoleh dari dataset TCGA Colon Cancer dan diproses melalui *variance threshold* serta ANOVA sebagai tahap *pre-filtering*, sebelum diterapkan FSACE. Fitur terpilih selanjutnya dievaluasi menggunakan model klasifikasi XGBoost dengan penanganan ketidakseimbangan kelas menggunakan SMOTE. Hasil eksperimen menunjukkan bahwa jumlah gen berhasil direduksi dari puluhan ribu menjadi enam gen informatif, dengan akurasi sebesar 79,82% dan nilai AUC sebesar 0,7975 dalam membedakan stadium awal dan lanjut kanker kolon. Hasil ini menunjukkan bahwa kombinasi FSACE dan XGBoost dapat digunakan sebagai pendekatan untuk klasifikasi kanker kolon berbasis data ekspresi gen berdimensi tinggi.

**Kata kunci:** Ekspresi Gen, FSACE, Kanker Usus Besar, Seleksi Fitur, XGBoost



# **FEATURE SELECTION OF GENE EXPRESSION USING FSACE FOR COLON CANCER PREDICTION**

Steven Lie

## **ABSTRACT**

*Gene expression data for colon cancer analysis are characterized by extremely high dimensionality and a limited number of samples, which often reduces classification performance and increases the risk of overfitting. Feature selection is therefore essential to identify informative genes. This study proposes a filter-based approach using Feature Selection based on Approximate Conditional Entropy (FSACE) to reduce the dimensionality of colon cancer gene expression data. RNA-Seq data from the TCGA Colon Cancer dataset were preprocessed using variance thresholding and ANOVA as a pre-filtering stage, followed by FSACE to select the most relevant genes. The selected features were then evaluated using an XGBoost classification model, with class imbalance handled through SMOTE. Experimental results show that the proposed approach successfully reduces tens of thousands of genes to a small subset of six informative genes, achieving an accuracy of 79.82 percent and an AUC of 0.7975 in distinguishing early and late-stage colon cancer. These findings indicate that the combination of FSACE and XGBoost can be used as an alternative approach for colon cancer classification based on high-dimensional gene expression data.*

**Keywords:** Colon cancer, Feature selection, FSACE, Gene Expression, XGBoost



## DAFTAR ISI

|                                                               |      |
|---------------------------------------------------------------|------|
| HALAMAN JUDUL                                                 | i    |
| PERNYATAAN TIDAK MELAKUKAN PLAGIAT                            | ii   |
| PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)      | iii  |
| HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN                       | v    |
| HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH                    | vi   |
| HALAMAN PERSEMBAHAN/MOTO                                      | vii  |
| KATA PENGANTAR                                                | viii |
| ABSTRAK                                                       | ix   |
| ABSTRACT                                                      | x    |
| DAFTAR ISI                                                    | xi   |
| DAFTAR TABEL                                                  | xii  |
| DAFTAR GAMBAR                                                 | xiii |
| DAFTAR KODE                                                   | xiv  |
| DAFTAR RUMUS                                                  | xv   |
| DAFTAR LAMPIRAN                                               | xvi  |
| BAB 1 PENDAHULUAN                                             | 1    |
| 1.1 Latar Belakang Masalah                                    | 1    |
| 1.2 Rumusan Masalah                                           | 2    |
| 1.3 Batasan Permasalahan                                      | 2    |
| 1.4 Tujuan Penelitian                                         | 3    |
| 1.5 Urgensi Penelitian                                        | 3    |
| 1.6 Luaran Penelitian                                         | 3    |
| 1.7 Manfaat Penelitian                                        | 3    |
| 1.8 Sistematika Penulisan                                     | 4    |
| BAB 2 LANDASAN TEORI                                          | 5    |
| 2.1 Tinjauan Teori                                            | 5    |
| 2.1.1 Kanker Usus Besar dan Staging                           | 5    |
| 2.1.2 ANOVA (Analysis of Variance)                            | 6    |
| 2.1.3 Feature Selection Using Approximate Conditional Entropy | 7    |
| 2.1.4 Model Klasifikasi XGBoost                               | 8    |
| BAB 3 METODOLOGI PENELITIAN                                   | 10   |
| 3.1 Studi Literatur                                           | 11   |
| 3.2 Dataset                                                   | 11   |
| 3.3 Pra-proses data                                           | 11   |
| 3.4 Seleksi Fitur                                             | 12   |
| 3.5 Solusi Data Tidak Seimbang                                | 12   |
| 3.6 Pembangunan Model                                         | 12   |
| BAB 4 HASIL DAN DISKUSI                                       | 14   |
| 4.1 Pra-proses Data                                           | 14   |
| 4.1.1 Penggabungan Data dan Penentuan Label                   | 14   |
| 4.1.2 Pembagian Data Latih dan Data Uji                       | 15   |
| 4.1.3 Variance Threshold                                      | 15   |
| 4.1.4 ANOVA sebagai Prapenyaringan Fitur                      | 16   |
| 4.2 Seleksi Fitur Menggunakan FSACE                           | 17   |
| 4.2.1 Algoritma Seleksi Fitur FSACE                           | 17   |
| 4.2.2 Hasil Seleksi Fitur Menggunakan FSACE                   | 19   |
| 4.3 Pembangunan Model Klasifikasi                             | 22   |
| 4.3.1 Penanganan Ketidakseimbangan Kelas dengan SMOTE         | 22   |
| 4.3.2 Pelatihan Model Klasifikasi Menggunakan XGBoost         | 22   |
| 4.3.3 Evaluasi Kinerja Model                                  | 23   |
| BAB 5 SIMPULAN DAN SARAN                                      | 27   |
| 5.1 Simpulan                                                  | 27   |
| 5.2 Saran                                                     | 27   |
| DAFTAR PUSTAKA                                                | 28   |

## DAFTAR TABEL

|            |                                                                           |    |
|------------|---------------------------------------------------------------------------|----|
| Tabel 2.1  | Pengelompokan Stadium Kanker Usus Besar menurut AJCC Edisi ke-8           | 6  |
| Tabel 4.1  | Karakteristik Matriks Fitur Ekspresi Gen . . . . .                        | 14 |
| Tabel 4.2  | Pemetaan Stadium AJCC ke Kelas Kanker Usus Besar . . . . .                | 15 |
| Tabel 4.3  | Reduksi Fitur Ekspresi Gen Menggunakan Variance Threshold . . . . .       | 16 |
| Tabel 4.4  | Jumlah Fitur Ekspresi Gen Sebelum dan Sesudah Pra-seleksi ANOVA . . . . . | 17 |
| Tabel 4.5  | Hasil Perhitungan Importance of Internal Attribute (IIA) . . . . .        | 20 |
| Tabel 4.6  | Hasil Seleksi Core Attributes Berdasarkan Nilai IIA . . . . .             | 21 |
| Tabel 4.7  | Fitur Tambahan Hasil Seleksi IEA . . . . .                                | 21 |
| Tabel 4.8  | Parameter Model XGBoost . . . . .                                         | 23 |
| Tabel 4.9  | Hasil Akurasi untuk masing-masing Threshold . . . . .                     | 24 |
| Tabel 4.10 | Hasil Evaluasi Klasifikasi Model XGBoost . . . . .                        | 25 |



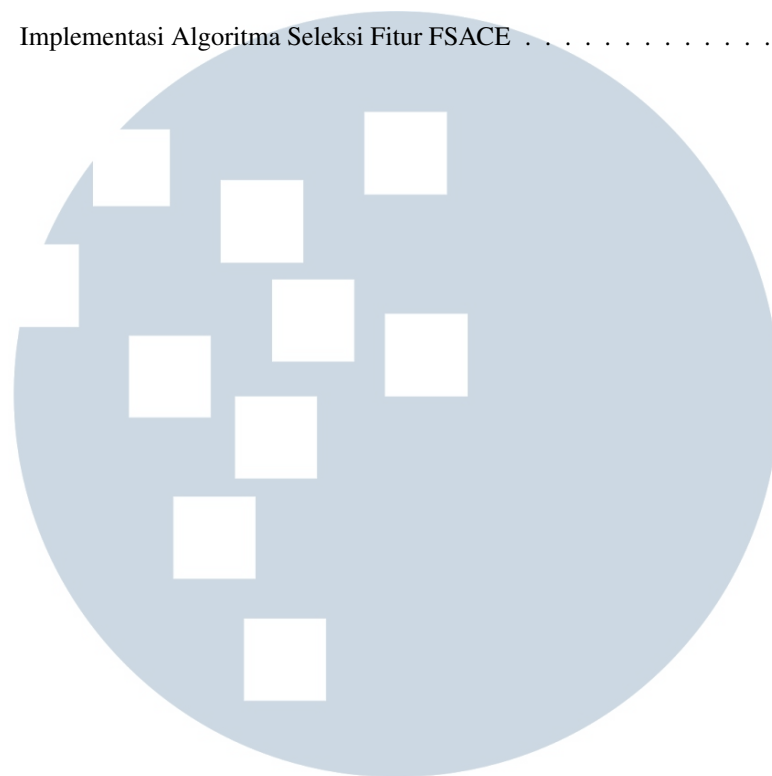
## DAFTAR GAMBAR

|            |                                                        |    |
|------------|--------------------------------------------------------|----|
| Gambar 3.1 | Pipeline Penelitian . . . . .                          | 10 |
| Gambar 4.1 | Confusion Matrix Model XGBoost pada Data Uji . . . . . | 25 |
| Gambar 4.2 | Kurva ROC Model XGBoost . . . . .                      | 26 |



## DAFTAR KODE

|          |                                                      |    |
|----------|------------------------------------------------------|----|
| Kode 4.1 | Implementasi Algoritma Seleksi Fitur FSACE . . . . . | 18 |
|----------|------------------------------------------------------|----|



UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR RUMUS

|           |                                                         |   |
|-----------|---------------------------------------------------------|---|
| Rumus 2.1 | Fuzzy Relation using Laplacian Kernel . . . . .         | 7 |
| Rumus 2.2 | Approximate Accuracy . . . . .                          | 7 |
| Rumus 2.3 | Approximate Conditional Entropy ( $H_{ace}$ ) . . . . . | 8 |
| Rumus 2.4 | IIA . . . . .                                           | 8 |
| Rumus 2.5 | IEA . . . . .                                           | 8 |
| Rumus 2.6 | XGBoost Prediction . . . . .                            | 8 |
| Rumus 2.7 | XGBoost Objective . . . . .                             | 9 |
| Rumus 2.8 | XGBoost Regularization . . . . .                        | 9 |





## DAFTAR LAMPIRAN

|             |                                                                                                                                             |    |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------|----|
| Lampiran 1  | Surat Pengantar PRO-STEP 01                                                                                                                 | 30 |
| Lampiran 2  | Kartu PRO-STEP 02                                                                                                                           | 31 |
| Lampiran 3  | Daily Task PRO-STEP 03                                                                                                                      | 32 |
| Lampiran 4  | Lembar Verifikasi PRO-STEP 04                                                                                                               | 43 |
| Lampiran 5  | Surat Penerimaan PRO-STEP 05                                                                                                                | 44 |
| Lampiran 6  | Form Bimbingan                                                                                                                              | 45 |
| Lampiran 7  | Pengecekan hasil <i>similarity</i> penelitian                                                                                               | 46 |
| Lampiran 8  | Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI)                                                                                    | 53 |
| Lampiran 9  | Research Paper                                                                                                                              | 54 |
| Lampiran 10 | Perjanjian Kerja Sama Antara Fakultas Kedokteran Brawijaya Dengan<br>Fakultas Teknik Informatika Universitas Multimedia Nusantara Tangerang | 64 |

