



**PENERAPAN ALGORITMA GENETIKA PADA KLASIFIKASI KANKER  
PAYUDARA MENGGUNAKAN XGBOOST**

**SKRIPSI**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh  
SITI NORFAIZAH  
NIM. 1811016120016**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU**

**JUNI 2025**

**SKRIPSI**

**PENERAPAN ALGORITMA GENETIKA PADA KLASIFIKASI KANKER  
PAYUDARA MENGGUNAKAN XGBOOST**

Oleh :

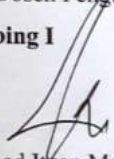
**SITI NORFAIZAH**

**1811016120016**

Telah dipertahankan di depan Dosen Penguji pada tanggal 30 Juni 2025

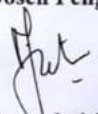
Susunan Dosen Penguji :

**Pembimbing I**



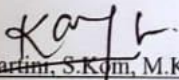
Muhammad Itqan Mazdadi S.Kom., M.Kom  
NIP. 199006122019031013

**Dosen Penguji I**



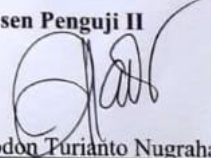
Fatma Indriani, S.T, M.I.T, Ph.D  
NIP. 198404202008122004

**Pembimbing II**



Dwi Kartini, S.Kom, M.Kom.  
NIP. 198704212012122003

**Dosen Penguji II**



Dodon Turianto Nugrahadi, S.Kom., M.Eng  
NIP. 198001122009121002

Panjabbaru, 30 Juni 2025

**Koordinator Program Studi Ilmu Komputer**



Dwi Kartini, S.Kom., M.Kom  
198704212012122003

## **PERNYATAAN**

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar keserjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Banjarbaru, 30 Juni 2025

Yang Menyatakan,



**SITI NORFAIZAH**  
**NIM. 1811016120016**

## ABSTRAK

### PENERAPAN ALGORITMA GENETIKA PADA KLASIFIKASI KANKER PAYUDARA MENGGUNAKAN XGBOOST

(Oleh: Siti Norfaizah; Pembimbing: Muhammad Itqan Mazdadi S.Kom., M.Kom. dan Dwi Kartini, S.Kom, M.Kom ; 2025; 50 halaman)

Kanker payudara merupakan salah satu penyakit mematikan yang tingkat keberhasilannya sangat bergantung pada deteksi dini yang akurat. Penelitian ini bertujuan untuk menganalisis kinerja algoritma XGBoost dalam klasifikasi kanker payudara serta mengevaluasi peningkatannya setelah dilakukan seleksi fitur menggunakan Algoritma Genetika (GA). Data yang digunakan adalah Breast Cancer Wisconsin (Diagnostic) dari UCI Machine Learning Repository yang terdiri dari 569 sampel dan 32 fitur. Proses penelitian meliputi preprocessing, pembentukan model XGBoost, dan penerapan seleksi fitur berbasis GA. Evaluasi model dilakukan menggunakan confusion matrix dengan teknik validasi silang 10-fold. Hasil penelitian menunjukkan bahwa XGBoost tanpa seleksi fitur menghasilkan akurasi 97,02%, recall 95,28%, precision 96,65%, dan F1-score 95,96%. Setelah dilakukan seleksi fitur menggunakan GA, jumlah fitur berkurang menjadi 22, dan kinerja model meningkat dengan akurasi 97,72%, recall 95,28%, precision 98,53%, dan F1-score 96,88%. Hasil ini membuktikan bahwa seleksi fitur menggunakan Algoritma Genetika tidak hanya menyederhanakan model, tetapi juga mampu meningkatkan akurasi dan efisiensi klasifikasi. Pendekatan ini sangat potensial untuk diterapkan dalam sistem pendukung keputusan medis guna meningkatkan ketepatan diagnosis kanker payudara.

**Kata Kunci:** Seleksi Fitur, *Genetic Algorithm* (GA), Klasifikasi, *XGBOOST*

## ABSTRACT

### ***APPLICATION OF GENETIC ALGORITHM IN BREAST CANCER CLASSIFICATION USING XGBOOST***

(By: Siti Norfaizah; Mentor: Muhammad Itqan Mazdadi S.Kom., M.Kom. and Dwi Kartini, S.Kom, M.Kom ; 2025; 50 pages)

*Breast cancer is one of the deadliest diseases, and its successful treatment highly depends on accurate early detection. This study aims to analyze the performance of the XGBoost algorithm in classifying breast cancer and evaluate its improvement after applying feature selection using a Genetic Algorithm (GA). The dataset used is the Breast Cancer Wisconsin (Diagnostic) from the UCI Machine Learning Repository, consisting of 569 samples and 32 features. The research process includes preprocessing, XGBoost model construction, and GA-based feature selection. Model evaluation was conducted using a confusion matrix with 10-fold cross-validation. The results show that XGBoost without feature selection achieved an accuracy of 97.02%, recall of 95.28%, precision of 96.65%, and an F1-score of 95.96%. After applying GA-based feature selection, the number of features was reduced to 22, and the model's performance improved with an accuracy of 97.72%, recall of 95.28%, precision of 98.53%, and an F1-score of 96.88%. These results indicate that feature selection using a Genetic Algorithm not only simplifies the model but also improves classification accuracy and efficiency. This approach shows strong potential for implementation in medical decision support systems to enhance the precision of breast cancer diagnosis.*

**Keywords:** *Feature Selection, Genetic Algorithm (GA), Classification, XGBOOST*

## PRAKATA

Puji syukur penulis panjatkan kepada Allah SWT karena atas berkat, rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **PENERAPAN ALGORITMA GENETIKA PADA KLASIFIKASI KANKER PAYUDARA MENGGUNAKAN XGBOOST** untuk memenuhi syarat dalam menyelesaikan pendidikan program S-1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa penulis panjatkan sholawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga *yaumul qiama*.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak – pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :

- (1) Keluarga yang selalu memberikan doa, semangat, bantuan dan dukungan dalam proses penyelesaian skripsi ini.
- (2) Seluruh dosen dan staf program studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama berkuliah di Ilmu Komputer yang sangat bermanfaat.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan skripsi ini masih jauh dari kesempurnaan, maka penulis mengharapkan bantuan berupa kritik dan saran yang membangun dari semua pihak. Semoga tulisan ini pula dapat memberikan manfaat bagi semua pihak untuk dapat menambah pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 30 Juni 2025



Penulis

## DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN .....	<b>Error! Bookmark not defined.</b>
PERNYATAAN.....	ii
ABSTRAK .....	iii
ABSTRACT .....	iv
PRAKATA .....	v
DAFTAR ISI.....	vi
DAFTAR TABEL .....	viii
DAFTAR GAMBAR .....	ix
DAFTAR LAMPIRAN.....	x
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian.....	2
1.4 Manfaat Penelitian.....	2
1.5 Batasan Masalah .....	3
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>4</b>
2.1 Kajian Terdahulu .....	4
2.2 Landasan Teori .....	8
2.2.1 Kanker Payudara.....	8
2.2.2 <i>Label Encoding</i> .....	9
2.2.3 Normalisasi Data.....	10
2.2.4 Seleksi Fitur .....	10
2.2.5 Algoritma Genetika.....	11
2.2.6 Klasifikasi .....	14
2.2.7 <i>Extreme Gradient Boosting</i> .....	15
2.2.8 <i>Cross Validation</i> .....	18
2.2.9 <i>Confusion Matrix</i> .....	19
<b>BAB III METODE PENELITIAN .....</b>	<b>21</b>

3.1	Alat Penelitian .....	21
3.2	Bahan Penelitian .....	21
3.3	Prosedur Penelitian .....	22
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>		<b>25</b>
4.1	Hasil.....	25
4.1.1	Pengumpulan Data.....	25
4.1.2	<i>Preprocessing</i> .....	27
4.1.3	Klasifikasi dan Evaluasi.....	29
4.2	Pembahasan .....	41
<b>BAB V PENUTUP.....</b>		<b>50</b>
5.1	Kesimpulan.....	50
5.1	Saran .....	50
DAFTAR PUSTAKA .....		xi
LAMPIRAN		

## DAFTAR TABEL

Tabel 1. Keaslian Penelitian.....	6
Tabel 2. Perancangan Penelitian .....	8
Tabel 3. <i>Confusion Matrix</i> .....	20
Tabel 4. Dataset Kanker Payudara .....	21
Tabel 5. Contoh Dataset Kanker Payudara .....	25
Tabel 6. Fitur Dataset <i>Breast Cancer Wisconsin (Diagnostic)</i> .....	26
Tabel 7. Hasil Normalisasi <i>Min-Max Normalization</i> .....	28
Tabel 8. Hasil Akurasi XGBoost .....	30
Tabel 9. Evaluasi Terbaik Klasifikasi XGBoost .....	33
Tabel 10. Nilai <i>Fitness</i> Hasil Kombinasi Fitur .....	35
Tabel 11. Fitur Terpilih pada Seleksi Fitur GA .....	36
Tabel 12. Hasil Akurasi XGBoost dengan Seleksi Fitur Algoritma Genetika.....	39
Tabel 13. Hasil Evaluasi Terbaik Klasifikasi XGBoost dengan Seleksi Fitur GA....	40
Tabel 14. Contoh Hitungan Manual Klasifikasi XGBoost .....	42
Tabel 15. Contoh Hitungan Manual Seleksi Fitur GA.....	44
Tabel 16. Analisa Hasil Perbandingan .....	48

## DAFTAR GAMBAR

Gambar 1. Pengkodean Biner .....	12
Gambar 2. One Point Crossover.....	12
Gambar 3. bit-flipping mutation .....	13
Gambar 4. Diagram Skema dari Algoritma XGBoost .....	15
Gambar 5. Konsep Dasar XGBoost .....	16
Gambar 6. Pembagian Data pada 10-Fold Cross Validation .....	19
Gambar 7. Alur Penelitian.....	23
Gambar 8. Hasil Confusion Matrix Klasifikasi XGBoost .....	33
Gambar 9. Grafik Evaluasi Terbaik Klasifikasi XGBoost Dengan Seleksi Fitur GA 39	
Gambar 10. Hasil Confusion Matrix Klasifikasi XGBoost dan Seleksi Fitur GA.....	40
Gambar 11. Analisa Perbandingan Hasil Dua Tahapan Penelitian.....	49

## DAFTAR LAMPIRAN

- Lampiran 1.** *Source Code Import Library*
- Lampiran 2.** *Source Code* Memisahkan Dataset Menjadi Data Dengan Label
- Lampiran 3.** *Source Code Pre-Processing* Normalisasi Data Dan *Label Encoder* Menjadi Boolean
- Lampiran 4.** *Source Code* Menampilkan Tabel Hasil Normalisasi
- Lampiran 5.** *Source Code* Parameter Xgboost
- Lampiran 6.** *Source Code* Implementasi *Genetic Algorithm*
- Lampiran 7.** *Source Code* Menampilkan Fitur Pilihan Hasil Implementasi *Genetic Algorithm*
- Lampiran 8.** *Source Code* Menampilkan Data Training Dengan Semua Fitur
- Lampiran 9.** *Source Code* Membuat *Function* Evaluasi
- Lampiran 10.** *Source Code* Mendapatkan Akurasi Sebelum Seleksi Fitur
- Lampiran 11.** *Source Code* Mendapatkan Akurasi Setelah Seleksi Fitur
- Lampiran 12.** *Source Code* Memasukkan Nilai Ke Dalam File Csv
- Lampiran 13.** *Source Code* Membuat Fungsi Untuk Gambar *Confusion Matrix*
- Lampiran 14.** *Source Code* Menampilkan Akurasi dan *Confusion Matrix*
- Lampiran 15.** Hasil Akurasi XGBoost dengan Seleksi Fitur GA
- Lampiran 16.** Nilai *Fitness* dari Kombinasi Fitur Seleksi GA