

**PERBANDINGAN ALGORITMA GENETIKA DAN ELIMINASI  
FITUR REKURSIF PADA DATA BERDIMENSI TINGGI**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Sistem Informasi



disusun oleh

**DIPA WIRANTANU**

**20.12.1695**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2024**

**PERBANDINGAN ALGORITMA GENETIKA DAN ELIMINASI  
FITUR REKURSIF PADA DATA BERDIMENSI TINGGI**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Sistem Informasi



disusun oleh

**DIPA WIRANTANU**

**20.12.1695**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA**

**YOGYAKARTA**

**2024**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**PERBANDINGAN ALGORITMA GENETIKA DAN ELIMINASI FITUR  
REKURSIF PADA DATA BERDIMENSI TINGGI**

yang disusun dan diajukan oleh

**Dipa Wirantanu**

**20.12.1695**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 7 Desember 2023

Dosen Pembimbing,



Yoga Pristyanto, S.Kom, M.Eng  
NIK. 190302412

**HALAMAN PENGESAHAN**

**SKRIPSI**

**PERBANDINGAN ALGORITMA GENETIKA DAN ELIMINASI FITUR  
REKURSIF PADA DATA BERDIMENSI TINGGI**

yang disusun dan diajukan oleh

**Dipa Wirantani**

**20.12.1695**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 18 Desember 2023

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Ali Mustopa, M. Kom**

**NIK. 190302192**

**Atik Nurmasani, S. Kom., M. Kom**

**NIK. 190302354**

**Yoga Pristyanto, S. Kom, M. Eng**

**NIK. 190302412**

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 18 Desember 2023

**DEKAN FAKULTAS ILMU KOMPUTER**



**Hanif Al Fatta, S.Kom., M.Kom., Ph.D.**

**NIK. 190302096**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Dipa Wirantanu  
NIM : 20.12.1695

Menyatakan bahwa Skripsi dengan judul berikut:

### **Perbandingan Algoritma Genetika dan Eliminasi Fitur Rekursif Pada Data Berdimensi Tinggi**

Dosen Pembimbing : Yoga Pristyanto, S.Kom. M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 18 Desember 2023.

Yang Menyatakan,



Dipa Wirantanu

## HALAMAN PERSEMBAHAN

Skripsi ini penulis persembahkan kepada kedua orang tua dan adik tercinta, dengan ketulusan doa, dan semangat mendukung setiap saat. Serta untuk beberapa teman kampus, dan untuk segenap dosen sekaligus rektor terimakasih banyak sudah membimbing selama tiga tahun.





## KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat, hidayat, serta kesempatan kepada penulis untuk menyelesaikan Tugas Akhir dalam rangka memenuhi persyaratan untuk mendapatkan gelar Sarjana Komputer dengan judul **“Perbandingan Algoritma Genetika dan Eliminasi Fitur Rekursif Pada Data Berdimensi Tinggi”**. Berkenaan dengan pembuatan Tugas Akhir tersebut, penulis menyampaikan terima kasih kepada yang terhormat:

1. Bapak Yoga Pristyanto, S.Kom, M.Eng, selaku Dosen Pembimbing utama yang telah memberikan bimbingan dan arahan selama proses penyusunan tugas akhir ini,
2. Bapak Ali Mustopa, M. Kom, ibu Atik Nurmasani, S. Kom., M. Kom selaku Dosen Penguji yang telah menguji selama proses sidang dari awal sampai selesai.
3. Bapak, Ibu, dan Adik tercinta yang memberikan dukungan serta doa sehingga Tugas Akhir ini bisa diselesaikan.
4. Beberapa teman penulis di Universitas Amikom Yogyakarta khususnya kelas SI 05 yang telah membantu dan mendukung selama pembuatan Tugas Akhir ini.

Dengan ini, penulisan Tugas Akhir ini dibuat, masih banyak kesalahan dan kekurangan dalam penulisan maupun isi sehingga segala kritik dan saran yang membangun dapat menyempurnakan Tugas Akhir serta dapat bermanfaat bagi penulis dan pembaca.

Yogyakarta, 31 Oktober 2023

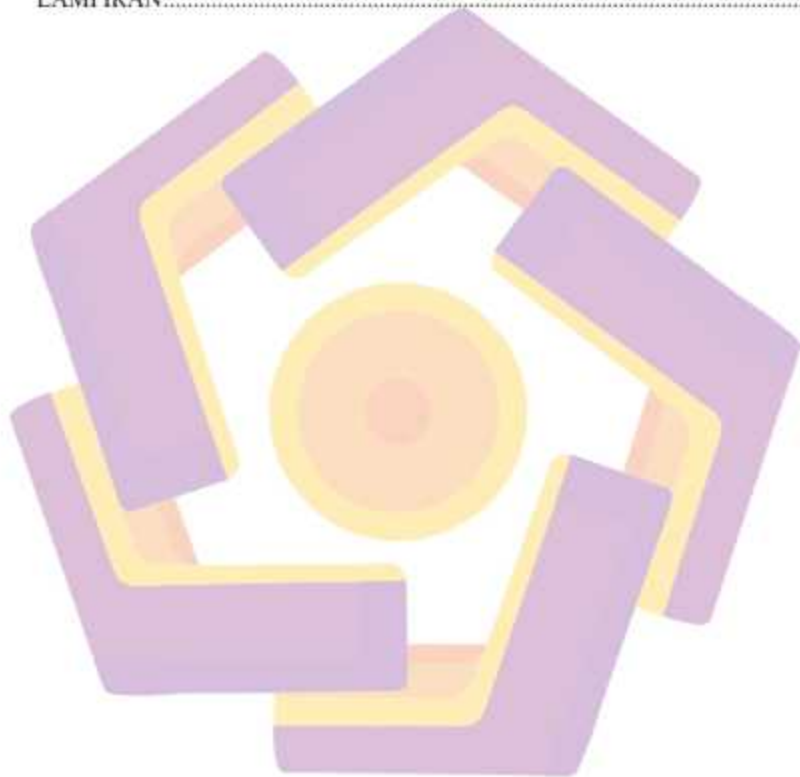
Penulis

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI .....	iv
HALAMAN PERSEMBAHAN .....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN.....	xi
DAFTAR LAMBANG DAN SINGKATAN .....	xii
DAFTAR ISTILAH .....	xiii
INTISARI .....	xiv
<i>ABSTRACT</i> .....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah .....	4
1.4 Tujuan Penelitian .....	5
1.5 Manfaat Penelitian .....	5
1.6 Sistematika Penulisan .....	5
BAB II TINJAUAN PUSTAKA .....	6
2.1 Studi Literatur.....	6
2.2 Dasar Teori .....	13
BAB III METODE PENELITIAN .....	22
3.1 Objek Penelitian.....	22
3.2 Alur Penelitian .....	22
3.3 Alat dan Bahan.....	25



BAB IV HASIL DAN PEMBAHASAN .....	28
BAB V PENUTUP .....	69
5.1 Kesimpulan .....	69
5.2 Saran .....	69
REFERENSI .....	70
LAMPIRAN.....	73



## DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian .....	9
Tabel 4.1 hasil fitur seleksi test size 0.1 .....	28
Tabel 4.2 hasil fitur seleksi test size 0.2 .....	29
Tabel 4.3 hasil fitur seleksi test size 0.3 .....	30
Tabel 4.4 test size 0.1 fitur seleksi GA .....	34
Tabel 4.5 test size 0.1 fitur seleksi RFE .....	35
Tabel 4.6 test size 0.2 fitur seleksi GA .....	36
Tabel 4.7 test size 0.2 fitur seleksi RFE .....	37
Tabel 4.8 test size 0.3 fitur seleksi GA .....	38
Tabel 4.9 test size 0.3 fitur seleksi RFE .....	39

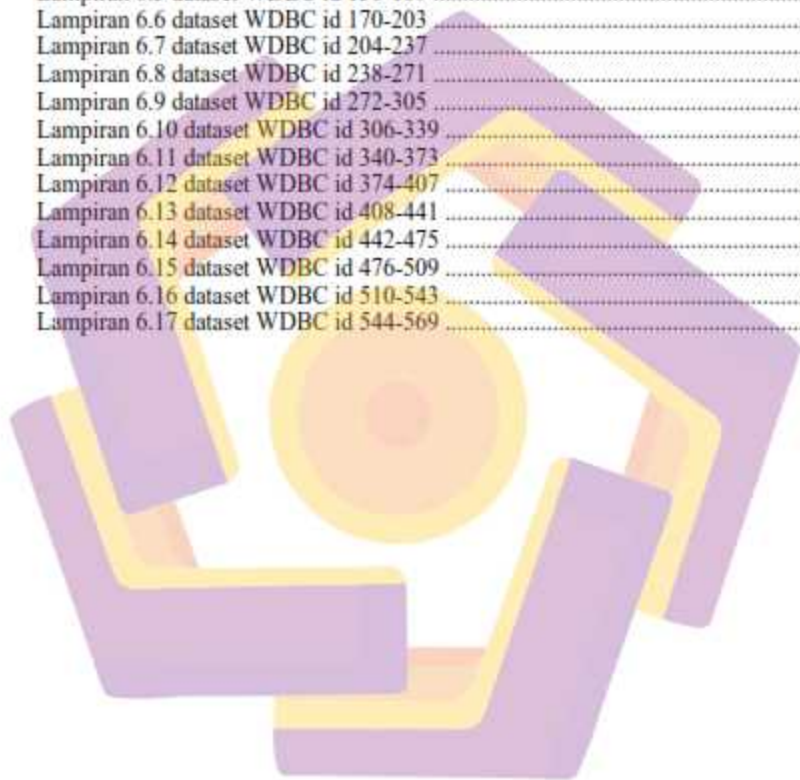


## DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian .....	22
Gambar 4.1 Coding test size dan random state .....	31
Gambar 4.2 Coding data split pada GA .....	32
Gambar 4.3 Coding data split pada RFE.....	32
Gambar 4.4 Model library algoritma .....	33
Gambar 4.5 Model library evaluasi .....	33
Gambar 4.6 halaman input CSV .....	40
Gambar 4.7 tampilan file sudah diupload pada halaman input CSV .....	41
Gambar 4.8 tampilan file belum diupload pada halaman classification .....	41
Gambar 4.9 halaman classification bagian menampilkan seluruh data .....	42
Gambar 4.10 halaman classification bagian memilih target (y) dan konversi.....	42
Gambar 4.11 halaman classification bagian set data split dan fitur seleksi RFE...	43
Gambar 4.12 halaman classification bagian set data split dan fitur seleksi GA ...	44
Gambar 4.13 tampilan file belum diupload pada halaman prediction .....	44
Gambar 4.14 tampilan belum uji klasifikasi pada fitur seleksi RFE .....	45
Gambar 4.15 tampilan belum uji klasifikasi pada fitur seleksi GA .....	45
Gambar 4.16 tampilan prediksi dengan menggunakan fitur seleksi RFE.....	46
Gambar 4.17 tampilan prediksi dengan menggunakan fitur seleksi GA .....	46
Gambar 4.18 Save model RFE.....	47
Gambar 4.19 Save model GA .....	47
Gambar 4.20 Read model RFE .....	47
Gambar 4.21 Read model GA .....	48
Gambar 4.22 Input peneliti test size 0.1 pada fitur seleksi GA .....	49
Gambar 4.23 Prediksi model test size 0.1 pada fitur seleksi GA.....	50
Gambar 4.24 Input peneliti test size 0.1 pada fitur seleksi RFE.....	51
Gambar 4.25 Prediksi model test size 0.1 pada fitur seleksi RFE .....	52
Gambar 4.26 Input peneliti test size 0.2 pada fitur seleksi GA .....	53
Gambar 4.27 Prediksi model test size 0.2 pada fitur seleksi GA.....	54
Gambar 4.28 Input peneliti test size 0.2 pada fitur seleksi RFE.....	55
Gambar 4.29 Prediksi model test size 0.2 pada fitur seleksi RFE .....	56
Gambar 4.30 Input peneliti test size 0.3 pada fitur seleksi GA .....	57
Gambar 4.31 Prediksi model test size 0.3 pada fitur seleksi GA.....	58
Gambar 4.32 Input peneliti test size 0.3 pada fitur seleksi RFE.....	59
Gambar 4.33 Prediksi model test size 0.3 pada fitur seleksi RFE .....	60

## DAFTAR LAMPIRAN

Lampiran 6.1 dataset WDBC id 1-33 .....	73
Lampiran 6.2 dataset WDBC id 34-67 .....	74
Lampiran 6.3 dataset WDBC id 68-101 .....	75
Lampiran 6.4 dataset WDBC id 102-135 .....	76
Lampiran 6.5 dataset WDBC id 136-169 .....	77
Lampiran 6.6 dataset WDBC id 170-203 .....	78
Lampiran 6.7 dataset WDBC id 204-237 .....	79
Lampiran 6.8 dataset WDBC id 238-271 .....	80
Lampiran 6.9 dataset WDBC id 272-305 .....	81
Lampiran 6.10 dataset WDBC id 306-339 .....	82
Lampiran 6.11 dataset WDBC id 340-373 .....	83
Lampiran 6.12 dataset WDBC id 374-407 .....	84
Lampiran 6.13 dataset WDBC id 408-441 .....	85
Lampiran 6.14 dataset WDBC id 442-475 .....	86
Lampiran 6.15 dataset WDBC id 476-509 .....	87
Lampiran 6.16 dataset WDBC id 510-543 .....	88
Lampiran 6.17 dataset WDBC id 544-569 .....	89



## DAFTAR LAMBANG DAN SINGKATAN



R	String
$M_j$	Data pengganti
$D_0$	Data pengganti
$\lambda$	Vektor kolom berdimensi n
OOB	Out Of Bag
TP	True Positif
TN	True Negatif
FP	False Positif
FN	False Negatif
RFE	Recursive Feature Elimination
GA	Genetic Algorithm
FS	Feature Selection
LR	Logistic Regression
KNN	K-Nearest Neighbors
SVM	Support Vector Machine
RF	Random Forest
WDBC	Wisconsin Diagnostic Breast Center
UCI	UC Irvine
B	Benign
M	Malignant
FNA	Fine Needle Aspirate
MSM-T	Multisurface Method-Tree

## DAFTAR ISTILAH

Vektor	Besaran yang mempunyai arah
String	Representasi urutan karakter atau teks
Kanker	Pertumbuhan sel yang tidak terkendali
Klasifikasi	Proses mengategorikan data ke dalam kelas
Prediksi	Proses memperkirakan nilai mendatang berdasar informasi yang tersedia
Kernel	Mengukur kemiripan antara pasangan titik data dalam ruang fitur asli
Estimator	Memperkirakan parameter yang ada dalam data
Selector	Pemilihan atribut yang paling penting dalam suatu dataset
Data split	Proses membagi menjadi dua atau lebih subset yang berbeda
Akurasi	Mengukur sejauh mana model klasifikasi memberikan prediksi yang benar secara keseluruhan
Presisi	Mengukur sejauh mana prediksi positif model adalah benar
Recall	Mengukur sejauh mana model mampu menemukan semua positif yang sebenarnya
F1 score	Untuk mempertimbangkan keseimbangan presisi dan recall
Fitting	Proses menyesuaikan model ke data
User Interface	Segala interaksi antara pengguna dan suatu aplikasi
Integrasi	Proses menghubungkan berbagai komponen
Fitur Seleksi	Proses memilih subset fitur yang paling relevan dan penting dari suatu dataset
Framework	Struktur kerja yang menyediakan fondasi untuk membangun perangkat lunak
Streamlit	Framework yang digunakan untuk analisis data atau demonstrasi konsep dengan cepat



## INTISARI

Saat ini big data pada suatu perusahaan digunakan sebagai *data analytic*. *Data analytic* pada suatu perusahaan mengelola arsip data yang besar sehingga memunculkan masalah dalam pemrosesan file untuk mempercepat pengembangan arsip perusahaan. Pemrosesan file tersebut melibatkan penggabungan informasi dari berbagai sumber dengan representasi yang berbeda sehingga memiliki keragaman dan kualitas data yang bervariasi dari satu sumber ke sumber lainnya. Dengan kata lain file data pada big data saat ini memiliki sejumlah fitur data yang banyak yang mampu memengaruhi efektivitas prediksi.

Dari masalah diatas, peneliti menyelesaikan masalah dengan membangun *machine learning* berbasis website untuk mengubah *data analytic* menjadi dunia kecerdasan dalam hal mempercepat pemrosesan file. Kemudian untuk mengatasi sejumlah fitur data yang banyak dengan menggunakan perbandingan metode antara fitur seleksi *Recursive Feature Elimination* (RFE) dan *Genetic Algorithm* (GA).

Hasil pada penelitian ini dibagi menjadi tiga test size yaitu pertama 0.1, model algoritma nilai evaluasi tertinggi akurasi, presisi, recall f1 score pada fitur seleksi GA berturut-turut adalah 98.25% LR,RF; 98.36% RF, 98.25% LR,RF; 98.26% RF, sedangkan pada RFE berturut-turut adalah 98.25% LR,SVM,AdaBoost; 98.36% LR,AdaBoost; 98.25% LR,SVM,AdaBoost; 98.26% LR,AdaBoost. Test size 0.2 model algoritma nilai evaluasi tertinggi akurasi, presisi, recall f1 score pada fitur seleksi GA berturut-turut adalah 97.37% LR; 97.40% LR, 97.37% LR; 97.38% LR, sedangkan pada RFE berturut-turut adalah 98.25% LR; 98.29% LR; 98.25% LR; 98.23% LR. Test size 0.3 model algoritma nilai evaluasi tertinggi akurasi, presisi, recall f1 score pada fitur seleksi GA berturut-turut adalah 96.49% KNN; 96.49% KNN, 96.49% KNN; 96.48% KNN, sedangkan pada RFE berturut-turut adalah 95.91% LR; 95.93% LR; 95.91% LR; 95.91% LR. Penelitian ini dapat digunakan untuk memprediksi pada data berdimensi tinggi. Dari hasil penelitian ini dapat digunakan oleh peneliti yang ada di daerah wisconsin, United States untuk memprediksi terkait kanker payudara.

**Kata kunci:** Big data, Algoritma Machine learning, Fitur seleksi, *Recursive Feature Elimination*, *Genetic Algorithm*

## ABSTRACT

Currently, big data in a company is used as data analytics. Data analytics in a company manages a large data archive, which raises problems in file processing to accelerate the development of company archives. The file processing involves combining information from various sources with different representations so that it has diversity and data quality that varies from one source to another. In other words, data files in big data today have a large number of data features that can affect the effectiveness of prediction.

From the above problems, researchers solve the problem by building web-based machine learning to turn analytic data into the world of intelligence in terms of accelerating file processing. Then to overcome a large number of data features by using a method comparison between Recursive Feature Elimination (RFE) selection features and Genetic Algorithm (GA).

The results in this study are divided into three test sizes, namely the first 0.1, the highest evaluation value algorithm model accuracy, precision, recall f1 score on the GA selection feature is 98.25% LR, RF; 98.36% RF, 98.25% LR, RF; 98.26% RF, while in RFE successively is 98.25% LR, SVM, AdaBoost; 98.36% LR, AdaBoost; 98.25% LR, SVM, AdaBoost; 98.26% LR, AdaBoost. Test size 0.2 algorithm model the highest evaluation values of accuracy, precision, recall f1 score on GA selection features are 97.37% LR; 97.40% LR, 97.37% LR; 97.38% LR, while on RFE they are 98.25% LR; 98.29% LR; 98.25% LR; 98.23% LR. Test size 0.3 algorithm model the highest evaluation value of accuracy, precision, recall f1 score on GA selection features are 96.49% KNN; 96.49% KNN, 96.49% KNN; 96.48% KNN, while in RFE they are 95.91% LR; 95.93% LR; 95.91% LR; 95.91% LR. This research can be used to predict on high-dimensional data. The results of this study can be used by researchers in the Wisconsin area, United States to predict breast cancer.

**Keyword:** Big data, Machine learning algorithm, Feature selection, Recursive Feature Elimination, Genetic Algorithm