

**OPTIMASI *RANDOM FOREST* DENGAN *GA* DAN *RFE* PADA *HIGH DIMENSIONAL DATA STUNTING***

**SKRIPSI**

**Diajukan Oleh :**

**Bima Satria**

**2011102441102**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR  
JULI 2024**

**OPTIMASI *RANDOM FOREST* DENGAN *GA* DAN *RFE* PADA *HIGH DIMENSIONAL DATA STUNTING***

**SKRIPSI**

Diajukan Sebagai Salah Satu Persyaratan Untuk Memperoleh Gelar  
Sarjana Komputer

**Diajukan Oleh :**  
**Bima Satria**  
**2011102441102**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR  
JULI 2024**

**LEMBAR PERSETUJUAN**  
**OPTIMASI *RANDOM FOREST* DENGAN *GA* DAN *RFE* PADA *HIGH***  
***DIMENSIONAL* DATA STUNTING**

**SKRIPSI**

Diajukan oleh:

Bima Satria  
2011102441102

Disetujui untuk diujikan Pada tanggal 4. Juli 2024

Pembimbing



Taghfirul Azhima Yoga Siswa, S.Kom, M.Kom  
NIDN. 1118038805

28/6/2024

Mengetahui,  
Koordinator Skripsi



Abdul Rahim, S.Kom, M.Cs.  
NIDN. 1115039601


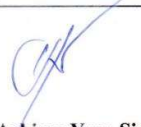
**LEMBAR PENGESAHAN**  
**OPTIMASI *RANDOM FOREST* DENGAN *GA* DAN *RFE* PADA *HIGH DIMENSIONAL* DATA STUNTING**

**SKRIPSI**

Diajukan oleh:

Bima Satria  
2011102441102

Diseminarkan dan Diujikan  
Pada tanggal 4.. Juli 2024

Penguji I	Penguji II
 <b><u>Wawan Joko Pranoto, S.Kom, M.Ti</u></b> NIDN. 1102057701	 <b><u>Taghfirul Azhima Yoga Siswa, S.Kom, M.Kom</u></b> NIDN. 1118038805

Mengetahui,

Ketua

Program Studi Teknik Informatika


**Arbansyah, S.Kom, M. TI**  
NIDN.1118019203

## PERNYATAAN KEASLIAN PENELITIAN

Saya yang bertanda tangan di bawah ini:

Nama : Bima Satria

NIM : 2011102441102

Program Studi : Teknik Informatika

Judul Penelitian : Optimasi *Random Forest* Dengan *GA* Dan *RFE* Pada *High Dimensional* Data Stunting

Menyatakan bahwa skripsi yang saya tulis ini benar-benar hasil karya saya sendiri, dan bukan merupakan hasil plagiasi/falsifikasi/fabrikasi baik sebagian atau seluruhnya.

Atas pernyataan ini, saya siap menanggung risiko atau sanksi yang dijatuhkan kepada saya apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam skripsi saya ini, atau klaim dari pihak lain terhadap keaslian karya sayaini

Samarinda, 27 April 2024

Yang membuat pernyataan



Bima Satria

NIM: 2011102441102

## ABSTRAK

Stunting adalah masalah kekurangan gizi kronis yang mengakibatkan terganggunya pertumbuhan anak, dengan dampak jangka panjang pada pertumbuhan fisik, perkembangan kognitif, dan produktivitas di masa dewasa. Di Indonesia, prevalensi stunting masih di atas batas WHO, mencapai 24,4% berdasarkan Studi Status Gizi Indonesia (SSGI) 2021, dan di Kota Samarinda prevalensinya mencapai 24,7% pada tahun 2021 dengan 1.402 balita teridentifikasi mengalami stunting. Mengatasi masalah ini memerlukan pendekatan berbasis data yang lebih terstruktur untuk memberikan intervensi yang tepat sasaran. Penelitian ini menggunakan data dari Dinas Kesehatan Kota Samarinda, mencakup 150.474 data stunting, dan melibatkan pengumpulan data, pembersihan data, seleksi fitur, serta penerapan model klasifikasi. Penelitian ini bertujuan untuk meningkatkan akurasi klasifikasi data stunting di Kota Samarinda tahun 2023 menggunakan algoritma Random Forest yang disempurnakan melalui teknik seleksi fitur Recursive Feature Elimination (RFE) dan optimasi Genetic Algorithm (GA). Hasil seleksi fitur menggunakan RFE menunjukkan bahwa fitur yang paling berpengaruh adalah Berat, ZS TB/U, ZS BB/U, dan BB/U. Penerapan RFE meningkatkan rata-rata akurasi model dari 91.91% menjadi 93.64%, sementara optimasi dengan GA meningkatkan rata-rata akurasi menjadi 98.39%. Akurasi pasti meningkat dari 94.23% (model dasar) menjadi 97.10% (dengan RFE) dan mencapai 99.70% (dengan RFE dan GA). Kombinasi RFE dan GA terbukti efektif dalam mengatasi kompleksitas data dan meningkatkan keandalan prediksi stunting. Penelitian ini memberikan kontribusi signifikan pada pengembangan teknik machine learning untuk analisis data berdimensi tinggi dalam bidang kesehatan, dan diharapkan menjadi dasar bagi program intervensi yang lebih efektif dalam menangani masalah stunting di Indonesia.

*Kata Kunci : Random Forest, Recursive Feature Elimination, Genetic Algorithm, Klasifikasi, High Dimensional*

## ABSTRACT

Stunting is a chronic malnutrition problem that disrupts children's growth, with long-term impacts on physical growth, cognitive development, and productivity in adulthood. In Indonesia, the prevalence of stunting is still above the WHO threshold, reaching 24.4% according to the 2021 Indonesian Nutritional Status Study (SSGI), and in Samarinda City, the prevalence reached 24.7% in 2021 with 1,402 toddlers identified as stunted. Addressing this problem requires a more structured data-driven approach to provide targeted interventions. This study uses data from the Samarinda City Health Office, encompassing 150,474 stunting data points, and involves data collection, data cleaning, feature selection, and classification model application. This study aims to improve the accuracy of stunting data classification in Samarinda City in 2023 using the Random Forest algorithm enhanced with Recursive Feature Elimination (RFE) feature selection techniques and Genetic Algorithm (GA) optimization. The feature selection results using RFE show that the most influential features are Weight, ZS TB/U, ZS BB/U, and BB/U. The application of RFE increased the model's average accuracy from 91.91% to 93.64%, while GA optimization further increased the average accuracy to 98.39%. The definite accuracy increased from 94.23% (baseline model) to 97.10% (with RFE) and reached 99.70% (with RFE and GA). The combination of RFE and GA has proven effective in tackling data complexity and improving the reliability of stunting predictions. This study significantly contributes to the development of machine learning techniques for high-dimensional data analysis in health and is expected to be the foundation for more effective intervention programs in addressing stunting issues in Indonesia.

*Keywords: Random Forest, Recursive Feature Elimination, Genetic Algorithm, Classification, High Dimensional*

## PRAKATA

Puji dan syukur kehadirat Tuhan Yang Maha Esa, karena dengan rahmat dan karunia-Nya, saya dapat menyelesaikan skripsi ini sebagai salah satu syarat untuk menyelesaikan program Sarjana di Fakultas Sains Dan Teknologi, Universitas Muhammadiyah Kalimantan Timur. Skripsi ini berjudul "*OPTIMASI RANDOM FOREST DENGAN GA DAN RFE PADA HIGH DIMENSIONAL DATA STUNTING*".

Penyusunan Skripsi ini tidak terlepas dari bantuan berbagai pihak, yang dengan tulus dan penuh kesabaran membimbing, membantu, serta memberikan motivasi kepada saya. Oleh karena itu, pada kesempatan ini, saya ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Mamaku tercinta beliau bernama Yusimayati dan keluarga yang selalu memberikan doa, dukungan moral, dan motivasi tanpa henti.
2. Bapak Taghfirul Azhima Yoga Siswa, S.Kom, M.Kom, selaku pembimbing utama, atas bimbingan, arahan, kritik, dan saran yang telah diberikan selama proses penyusunan proposal ini.
3. Bapak Wawan Joko Pranoto, S.Kom, M.Ti, selaku penguji utama, atas masukan, kritik, dan sarannya selama proses penyusunan skripsi ini.
4. Teman-teman Satu RTA yang selalu memberikan semangat dan dukungan selama proses penyusunan proposal ini.
5. Muhammad Ghalib Ramadhan, S.Kom dan Ihsan Magribi, S.Kom yang selalu membantu saya mengerjakan skripsi ini hingga selesai.

Saya menyadari bahwa proposal skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat saya harapkan untuk perbaikan. Semoga proposal skripsi ini dapat bermanfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang Informatika dan bagi siapa saja yang berkepentingan.

Samarinda, 27 April 2024  
Penyusun,

Bima Satria  
NIM: 2011102441102



# DAFTAR ISI

	<b>Halaman</b>
Halaman Judul .....	ii
Halaman Persetujuan .....	iii
Halaman Pengesahan .....	iv
Pernyataan Keaslian Penelitian .....	v
Abstrak .....	vi
Abstract .....	vii
Prakata .....	viii
Daftar isi .....	ix
Daftar Tabel .....	xi
Daftar Gambar .....	xii
Daftar Lampiran .....	xiii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Tujuan Penelitian .....	2
1.4 Manfaat Penelitian .....	3
1.5 Batasan Masalah .....	3
BAB II METODE PENELITIAN .....	4
2.1 Objek Penelitian .....	4
2.2 Teknik Analisis Data .....	4
2.2.1 Pengumpulan Data .....	5
2.2.2 Data <i>Pre-Procesing</i> .....	5
2.2.3 <i>K-Fold Cross Validation</i> .....	7
2.2.4 Pembuatan Model .....	8
2.2.5 Evaluasi .....	14
BAB III HASIL DAN PEMBAHASAN .....	15
3.1 Hasil .....	15
3.2 Data Penelitian .....	15
3.3 Seleksi Dan Integrasi Data .....	18
3.3.1 Data <i>Cleaning</i> .....	18
3.3.2 Data <i>Transformation</i> .....	19

3.4 Hasil Pemodelan Setiap Algoritma.....	21
3.4.1 <i>Random Forest</i> .....	21
3.4.2 <i>Random Forest</i> Dengan Seleksi Fitur <i>Recursive Feature Elimination(RFE)</i> .....	24
3.4.3 <i>Random Forest</i> Dengan Seleksi Fitur <i>RFE</i> Dan Optimasi <i>Genetic Algorithm(GA)</i> .....	27
3.4.4 Perbandingan Hasil Akurasi .....	30
3.5 Pembahasan.....	31
BAB IV KESIMPULAN DAN SARAN.....	34
4.1 Kesimpulan.....	34
4.2 Saran.....	34
DAFTAR RUJUKAN.....	36
DAFTAR RIWAYAT HIDUP .....	39
LAMPIRAN .....	40

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
Tabel 2. 1 Atribut Data Stunting Dinas Kesehatan Kota Samarinda .....	5
Tabel 2. 2 Data Selection.....	6
Tabel 2. 3 Penjelasan Parameter Kode Pembagian Dataset Dan Perulangan .....	8
Tabel 2. 4 Parameter Model Algoritma <i>Random Forest</i> .....	11
Tabel 2. 5 Parameter Seleksi Fitur <i>RFE</i> .....	12
Tabel 2. 6 Parameter Optimasi <i>GA</i> .....	13
Tabel 2. 7 <i>Confusion Matrix</i> .....	14
Tabel 3. 1 Dataset Stunting Kota Samarinda Tahun 2023 .....	16
Tabel 3. 2 Hasil Seleksi dan Integrasi Data .....	18
Tabel 3. 3 Data Setelah <i>Cleaning</i> .....	18
Tabel 3. 4 Data Sebelum di <i>Transformation</i> .....	20
Tabel 3. 5 Data Setelah di <i>Transformation</i> .....	20
Tabel 3. 6 Hasil Seleksi Fitur <i>RFE</i> .....	21
Tabel 3. 7 Hasil Akurasi <i>Random Forest</i> .....	22
Tabel 3. 8 Penjelasan <i>Random Forest K-Fold = 10</i> .....	22
Tabel 3. 9 <i>Confusion Matrix Random Forest</i> .....	23
Tabel 3. 10 Hasil Akurasi <i>Random Forest + RFE</i> .....	24
Tabel 3. 11 Penjelasan <i>RF + RFE K-Fold = 10</i> .....	25
Tabel 3. 12 <i>Confusion Matrix RF + RFE</i> .....	26
Tabel 3. 13 Hasil Akurasi <i>Random Forest + RFE + GA</i> .....	27
Tabel 3. 14 Penjelasan <i>RF + RFE + GA K-Fold = 10</i> .....	28
Tabel 3. 15 <i>Confusion Matrix RF + RFE + GA</i> .....	29
Tabel 3. 16 Perbandingan Hasil Akurasi Sesudah Penggunaan <i>RFE &amp; GA</i> .....	30
Tabel 3. 17 Perbandingan Hasil Rata-Rata Akurasi Sesudah Penggunaan <i>RFE &amp; GA</i> .....	31
Tabel 3. 18 Perbandingan <i>RFE</i> Dengan Penelitian lain.....	32
Tabel 3. 18 Perbandingan <i>GA</i> Dengan Penelitian lain .....	33

## DAFTAR GAMBAR

Gambar	Halaman
Gambar 2.1 Alur Penelitian .....	4
Gambar 2.2 Kode <i>Cleaning</i> Data Kosong .....	6
Gambar 2.3 Kode <i>Data Transformation</i> .....	7
Gambar 2.4 Kode <i>K-Fold Cross Validation</i> .....	7
Gambar 2.5 Kode Data Pengukuran Terbaru.....	8
Gambar 2.6 Kode <i>Random Forest</i> .....	9
Gambar 2.7 Kode Model Algoritma <i>Random Forest</i> .....	10
Gambar 2.8 Kode Seleksi Fitur <i>RFE</i> .....	12
Gambar 2.9 Kode Optimasi <i>GA</i> .....	13
Gambar 3.1 Dataset Setelah Tahap <i>Cleaning</i> .....	19
Gambar 3.2 Data Setelah diubah kategorinya .....	21
Gambar 3.3 <i>Confusion Matrix Random Forest</i> .....	24
Gambar 3.4 <i>Confusion Matrix RF + RFE</i> .....	27
Gambar 3.5 <i>Confusion Matrix RF + RFE + GA</i> .....	29

## DAFTAR LAMPIRAN

<b>Lampiran</b>	<b>Halaman</b>
Lampiran 1 Surat Permohonan Pengambilan Data.....	41
Lampiran 2 Tampilan Dataset Stunting Kota Samarinda Tahun 2023.....	42
Lampiran 3 Laporan Bimbingan .....	43
Lampiran 4 Kode Klasifikasi Algoritma Random Forest + RFE + GA .....	45
Lampiran 5 Turnitin Skripsi .....	55