

**PERBAIKAN AKURASI *RANDOM FOREST* DENGAN *ANOVA* DAN  
*SMOTE* PADA KLASIFIKASI DATA STUNTING**

**SKRIPSI**

**Diajukan Oleh :  
Ari Ahmad Dhani  
2011102441090**



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

**PERBAIKAN AKURASI *RANDOM FOREST* DENGAN *ANOVA* DAN  
*SMOTE* PADA KLASIFIKASI DATA STUNTING**

**SKRIPSI**

Diajukan Sebagai Salah Satu Persyaratan Untuk Memperoleh Gelar  
Sarjana Komputer Fakultas Sains Dan Teknologi Universitas  
Muhammadiyah Kalimantan Timur

**Diajukan Oleh :  
Ari Ahmad Dhani  
2011102441090**



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

**LEMBAR PERSETUJUAN**

**PERBAIKAN AKURASI RANDOM FOREST DENGAN ANOVA DAN  
SMOTE PADA KLASIFIKASI DATA STUNTING**

**SKRIPSI**

**Diajukan oleh:**

**Ari Ahmad Dhani  
2011102441090**

**Disetujui untuk diujikan Pada tanggal 26 Juni 2024**

**Pembimbing**



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

**Mengetahui,  
Koordinator Skripsi**



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

**LEMBAR PENGESAHAN**



**PERBAIKAN AKURASI RANDOM FOREST DENGAN ANOVA DAN  
SMOTE PADA KLASIFIKASI DATA STUNTING**

**SKRIPSI**

**Diajukan oleh:**

**Ari Ahmad Dhani  
2011102441090**

**Diseminarkan dan Diujikan  
Pada tanggal 04 Juli 2024**

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

**Mengetahui,**

**Ketua**

**Program Studi Teknik Informatika**



  
Agus Syah, S.Kom, M.TI  
NIDN.1118019203

## PERNYATAAN KEASLIAN PENELITIAN

Saya yang bertanda tangan di bawah ini:

Nama : Ari Ahmad Dhani  
NIM : 2011102441090  
Program Studi : Teknik Informatika  
Judul Penelitian : Perbaikan Akurasi *Random Forest* Dengan *Anova* Dan *Smote* pada Klasifikasi Data Stunting

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

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, 11 Juli 2024

Yang membuat pernyataan



Ari Ahmad Dhani  
NIM: 2011102441090

## ABSTRAK

Stunting terus menjadi isu kesehatan masyarakat yang kritis di Indonesia, khususnya di Kota Samarinda yang mencatat prevalensi sebesar 25,3% pada tahun 2022, menjadi yang tertinggi kedua di Provinsi Kalimantan Timur. Di tengah prioritas nasional untuk riset 2020-2024, penggunaan data mining untuk klasifikasi stunting memperlihatkan potensi yang signifikan namun tetap menghadapi tantangan dalam menangani data berdimensi tinggi dan ketidakseimbangan kelas. Penelitian ini bertujuan untuk meningkatkan akurasi klasifikasi stunting menggunakan metode *Random Forest* yang diintegrasikan dengan seleksi fitur *ANOVA* dan teknik *SMOTE* untuk menyeimbangkan kelas. Data yang digunakan dalam penelitian ini bersumber dari Dinas Kesehatan Kota Samarinda, meliputi 26 puskesmas dengan 21 atribut dan total 150.466 record. Teknik validasi yang dipakai adalah cross-validation k-fold=10. Hasil menunjukkan peningkatan akurasi dari 98,83% menjadi 99,77% naik sebesar 0,94% setelah penerapan seleksi fitur *ANOVA*. Fitur ZS TB/U, ZS BB/U, dan BB/U diidentifikasi sebagai yang paling berpengaruh. Peningkatan ini menunjukkan efektivitas integrasi metode dalam mengatasi masalah stunting pada dataset yang kompleks dan tidak seimbang, ini diharapkan dapat mendukung kebijakan dan intervensi kesehatan lebih lanjut di kawasan tersebut.

Kata kunci: Klasifikasi, *Random Forest*, *ANOVA*, *SMOTE*, *High Dimensional*

## ABSTRACT

*Stunting continues to be a critical public health issue in Indonesia, particularly in Samarinda City, which recorded a prevalence of 25.3% in 2022, the second highest in East Kalimantan Province. Amidst the national research priorities for 2020-2024, the use of data mining for stunting classification shows significant potential but still faces challenges in handling high-dimensional data and class imbalance. This study aims to improve stunting classification accuracy using the Random Forest method integrated with ANOVA feature selection and SMOTE technique for class balancing. The data used in this study were sourced from the Samarinda City Health Office, encompassing 26 health centers with 21 attributes and a total of 150,466 records. The validation technique used is 10-fold cross-validation. The results show an accuracy increase from 98.83% to 99.77%, an increase of 0.94%, after applying ANOVA feature selection. The features ZS TB/U, ZS BB/U, and BB/U were identified as the most influential. This increase demonstrates the effectiveness of the method integration in addressing the stunting problem in complex and imbalanced datasets and is expected to support further health policies and interventions in the area.*

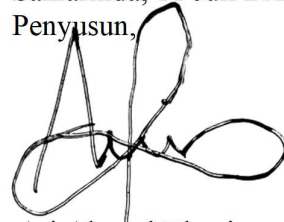
*Keywords: Classification, Random Forest, ANOVA, SMOTE, High Dimensional*

## PRAKATA

Alhamdulillah, puji syukur kehadirat Allah SWT yang telah melimpahkan Rahmat dan Karunia-Nya, sehingga peneliti dapat menyelesaikan skripsi dengan judul “Perbaikan Akurasi *RANDOM FOREST* Dengan *ANOVA* Dan *SMOTE* Pada Klasifikasi Data Stunting”. Dalam penyusunan skripsi ini, peneliti banyak mendapatkan bantuan dari beberapa pihak. Oleh karena itu, pada kesempatan ini peneliti ingin mengucapkan terima kasih kepada:

1. Bapak Sukari, Ibu Rumini, Kakak, dan Adik tercinta yang selalu memberikan doa serta dukungan kepada penulis.
2. Dr. Muhammad Musiyam, M.T, selaku Rektor Universitas Muhammadiyah Kalimantan Timur.
3. Bapak Arbansyah, S.Kom., M.TI selaku ketua Program Studi S1 Teknik Informatika.
4. Bapak Taghfirul Azhima Yoga Siswa, S.Kom, M.Kom selaku Dosen pembimbing yang telah memberikan wawasan dan arahan dalam pengerjaan skripsi ini.
5. Bapak Wawan Joko Pranoto, S.Kom, M.Ti selaku Dosen Penguji yang telah memberikan masukan dan arahan dalam revisi skripsi ini.
6. Seluruh Bapak dan Ibu Dosen Program Studi Teknik Informatika Universitas Muhammadiyah Kalimantan Timur dan 1 staff tendik yang penulis banggakan dan hormati.
7. Perpustakaan Daerah, Kota Samarinda, dan Perpustakaan Universitas Muhammadiyah Kalimantan Timur.
8. Tim RTA yang selalu siap membantu dalam pengerjaan skripsi ini
9. Tim GC Elite yaitu Renal, Bulan, Vito ttp semangat dan semoga dipermudah dalam meraih gelar S.Kom.
10. Pemilik Nim 2011102431071 terima kasih telah menjadi salah satu sosok penting baik dalam perkuliahan sampai ditahap ini, Telah berkontribusi banyak dalam penulisan skripsi ini, meluangkan waktu, pikiran, dan pengetahuan. Terima kasih karena sudah menjadi bagian dalam menuju gelar S.Kom semoga juga dipermudah dalam menuju gelar S.M dan semoga bisa menjadi Pribadi yang lebih baik lagi.

Samarinda, 11 Juli 2024  
Penyusun,



Ari Ahmad Dhani  
NIM: 2011102441090



## DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN PENELITIAN.....	v
ABSTRAK.....	vi
<i>ABSTRACT</i> .....	<i>vii</i>
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.....	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	3
BAB II METODOLOGI PENELITIAN.....	4
2.1 Objek Penelitian.....	4
2.2 Teknik Analisis Data.....	4
2.2.1 Pengumpulan data.....	5
2.2.2 <i>Data Pre-Processing</i> .....	5
2.2.3 Pembagian data.....	9
2.2.4 Pembuatan Model.....	10
2.2.5 Evaluasi.....	15
BAB III HASIL DAN PEMBAHASAN.....	16
3.1 Hasil Penelitian.....	16
3.1.1 <i>Data Selection</i> .....	16

3.1.2	<i>Data Cleaning</i> .....	16
3.1.3	<i>Data Transformation</i> .....	18
3.1.4	<i>Data Balancing</i> .....	18
3.1.5	Permodelan <i>Random forest</i> Tanpa <i>ANOVA</i> .....	20
3.1.6	Permodelan <i>Random Forest</i> Dengan <i>ANOVA</i> .....	21
3.1.7	Perbandingan Hasil.....	23
3.2	Pembahasan.....	24
BAB IV KESIMPULAN DAN SARAN .....		26
4.1	Kesimpulan.....	26
4.2	Saran.....	26
DAFTAR RUJUKAN.....		27
LAMPIRAN.....		29
RIWAYAT HIDUP.....		33

## DAFTAR TABEL

Tabel	Halaman
2.1 Atribut Data Dinas Kesehatan Kota Samarinda .....	5
2.2 Data <i>Selection</i> .....	6
2.3 Parameter Kode Menampilkan Nilai Kosong .....	6
2.4 Parameter Kode Menghapus Data Terduplikasi Dan Data Kosong .....	7
2.5 Parameter Kode Tranformasi .....	8
2.6 Parameter Kode Data Balancing <i>SMOTE</i> .....	9
2.7 <i>K-fold Cross Validation</i> .....	10
2.8 Parameter Pembagian Data Latih dan Data Uji .....	11
2.9 Parameter Permodelan Algoritma <i>Random Forest</i> .....	12
2.10 Parameter Hasil Skor Dari Setiap <i>Fold</i> .....	13
2.11 Parameter Seleksi Fitur <i>ANOVA</i> .....	15
3.1 Data <i>Selection</i> .....	16
3.2 Dataset sebelum dibersihkan.....	16
3.3 Dataset setelah dibersihkan .....	17
3.4 Dataset sebelum ditransformasi .....	18
3.5 Dataset Setelah ditransformasi.....	18
3.6 Hasil Pengujian model <i>Random Forest</i> .....	20
3.7 Rata Rata Akurasi Pengujian Model Random Forest.....	20
3.8 <i>Confusion Matrix</i> .....	20
3.9 Hasil Perengkingan <i>ANOVA</i> .....	22
3.10 Hasil Pengujian Model <i>Random Forest</i> .....	22
3.11 Rata Rata Akurasi pengujian model <i>Random Forest</i> .....	22
3.12 <i>Confusion Matrix</i> .....	23
3.13 Perbandingan Hasil Akurasi Pengujian <i>Random Forest</i> .....	24
3.14 Perbandingan Hasil Akurasi Rata-Rata <i>Random Forest</i> .....	24
3.15 Persamaan Hasil Seleksi Fitur Dengan Penelitian Lain .....	25

## DAFTAR GAMBAR

Gambar	Halaman
2.1 Alur Penelitian.....	4
2.2 Kode Menampilkan Nilai Kosong.....	6
2.3 Kode Menghapus Data Terduplikasi Dan Data Kosong .....	7
2.4 Kode Tranformasi Data Pada Kolom Atribut.....	8
2.5 Kode Tranformasi Data Pada Kolom Label .....	8
2.6 Kode Data <i>Balancing SMOTE</i> .....	9
2.7 Kode Inisialisasi Model <i>Random Forest</i> .....	11
2.8 Kode Pembagian Data Latih dan Data Uji .....	11
2.9 Kode Permodelan Algoritma <i>Random Forest</i> .....	12
2.10 Kode Hasil skor dari setiap <i>Fold</i> .....	13
2.11 Kode Seleksi Fitur <i>ANOVA</i> .....	14
3.1 Jumlah Nilai Kosong Tiap Kolom Sebelum Pembersihan .....	17
3.2 Jumlah Nilai Kosong Tiap Kolom Setelah Pembersihan .....	17
3.3 Jumlah Kelas Sebelum Data <i>Balancing</i> .....	19
3.4 Jumlah Kelas Sesudah Data <i>Balancing</i> .....	19
3.5 <i>Confusion Matrix Random Forest</i> .....	21
3.7 <i>Confusion Matrix Random Forest</i> .....	23

## DAFTAR LAMPIRAN

### Lampiran

Tampilan Dataset Stunting Kota Samarinda .....	29
Surat Pengantar Pengambilan Data .....	30
Kartu Bimbingan .....	31