

BAB III

HASIL DAN PEMBAHASAN

3.1 Hasil

Pada bagian ini, akan disajikan hasil analisis dan pemodelan data stunting di Kota Samarinda tahun 2023 yang telah dilakukan menggunakan berbagai metode yang berbeda. Penelitian ini mencakup serangkaian tahapan, mulai dari proses seleksi dan integrasi data, pembersihan data untuk memastikan kualitas dan konsistensi, hingga transformasi data agar lebih sesuai dengan kebutuhan algoritma pemodelan. Setiap tahap ini dirancang untuk meningkatkan akurasi dan efisiensi dari model yang akan digunakan.

Pada tahap pemodelan, berbagai teknik klasifikasi diterapkan untuk mengevaluasi performa masing-masing metode. Tahap awal dimulai dengan penerapan metode *Random Forest* dasar untuk mendapatkan gambaran awal dari hasil klasifikasi. Selanjutnya, dilakukan penambahan teknik seleksi fitur menggunakan *Recursive Feature Elimination (RFE)* untuk memilih fitur-fitur yang paling relevan dan signifikan dalam proses klasifikasi. Tahap terakhir melibatkan optimasi model menggunakan *Genetic Algorithm (GA)* untuk menemukan konfigurasi hyperparameter yang optimal, sehingga dapat meningkatkan akurasi dan kinerja model secara keseluruhan.

Hasil dari setiap tahap pemodelan ini akan dibahas secara rinci, termasuk analisis terhadap perubahan akurasi yang diperoleh pada setiap langkah. Selain itu, analisis perbandingan hasil akurasi dari masing-masing metode juga akan disajikan untuk memberikan gambaran yang komprehensif tentang kinerja masing-masing teknik dalam mengklasifikasikan data stunting di Kota Samarinda. Dengan demikian, dapat diketahui seberapa besar peningkatan performa yang diperoleh dari setiap tahapan seleksi fitur dan optimasi yang dilakukan, serta implikasi dari hasil tersebut terhadap upaya penanggulangan stunting di Kota Samarinda.

3.2 Data Penelitian

Penelitian ini menggunakan data stunting yang didapatkan dari Dinas Kesehatan Kota Samarinda. Data yang digunakan mencakup periode bulan Januari 2023 hingga Desember 2023 dengan total sebanyak 150.474 data. Data stunting yang diperoleh dari Dinas Kesehatan Samarinda memiliki 28 atribut yang penting dan mendetail. Atribut-atribut tersebut antara lain meliputi Nama, Jenis Kelamin (JK), Berat Badan saat Lahir (BB Lahir), Tinggi Badan saat Lahir (TB Lahir), Provinsi (Prov), Kabupaten/Kota (Kab/Kota), Kecamatan (Kec), Puskesmas, Desa/Kelurahan (Desa/Kel), Posyandu, Rukun Tetangga (RT), Rukun Warga (RW), Usia Saat Pengukuran (Usia Saat Ukur), Tanggal Pengukuran, Berat Badan (Berat), Tinggi Badan (Tinggi), Lingkar Lengan Atas (LiLA), Berat Badan per Usia (BB/U), Z-Score Berat Badan per Usia (ZS BB/U), Tinggi Badan per Usia (TB/U), Z-Score Tinggi Badan per Usia (ZS TB/U), Berat Badan per Tinggi Badan (BB/TB), Z-Score Berat Badan per Tinggi Badan (ZS BB/TB), Kenaikan Berat Badan (Naik Berat Badan), Pemberian Makanan Tambahan yang diterima dalam kilogram (PMT Diterima (kg)), Jumlah Vitamin A yang diterima (Jml Vit A), Kuesioner Praktek Sehari-hari Pengasuhan (KPSP), dan Kartu Ibu dan Anak (KIA).

Tabel 3.1 Dataset Stunting Kota Samarinda Tahun 2023

NO	1	2	3	:	150472	150473	15074
Nama	Dimas Aditya	Siti Aisyah	M Al Fatih	:	Afizah Khairina	M Arsyah Kholif	Muhamad Iqbal
JK	L	P	L	:	P	P	L
BB Lahir	3.5	3	2.8	:	2.5	3	2.9
TB Lahir	49	48	49	:	45	49	49
Provinsi	Kalimantan Timur	Kalimantan Timur	Kalimantan Timur	:	Kalimantan Timur	Kalimantan Timur	Kalimantan Timur
Kab/Kota	Samarinda	Samarinda	Samarinda	:	Samarinda	Samarinda	Samarinda
Kecamatan	Sungai Pinang	Sungai Pinang	Sungai Pinang	:	Samarinda Seberang	Samarinda Ulu	Samarinda Ilir
Puskesmas	Remaja	Remaja	Remaja	:	Mangkupalas	Juanda	Sidomulyo
Desa/Kel	Temindung Permai	Temindung Permai	Temindung Permai	:	Tenun Samarinda	Air Hitam	Sungai Dama
Posyandu	Pulau Indah	Pulau Indah	Pulau Indah	:	Balo Negara	Mekar Sejahtera	Ramania
RT	-	34	-	:	12	32	-
RW	-	-	-	:	-		
Usia Saat Ukur	0 Tahun - 11 Bulan - 11 Hari	4 Tahun - 0 Bulan - 16 Hari	0 Tahun - 7 Bulan - 7 Hari	:	0 Tahun - 0 Bulan - 0 Hari	0 Tahun - 0 Bulan - 0 Hari	0 Tahun - 0 Bulan - 0 Hari
Tanggal Pengukuran	2023-01-02	2023-01-02	2023-01-02	:	2023-12-09	2023-12-19	2023-12-29
Berat	9.1	12	8.1	:	2.5	3	2.9
Tinggi	74.5	94	69	:	45	49	49
LiLA	0	0	0	:	-	-	-
BB/U	Berat Badan Normal	Kurang	Berat Badan Normal	:	Kurang	Berat Badan Normal	Kurang
ZS BB/U	-0.39	-2.25	-0.53	:	-2.03	-1.56	-2.97
TB/U	Normal	Pendek	Normal	:	Pendek	Normal	Pendek
ZS TB/U	-0.21	-2.09	-0.65	:	-2.63	-1.3	-2.48

BB/TB	Gizi Baik	Gizi Baik	Gizi Baik	:	Gizi Baik	Gizi Baik	Gizi Baik
ZS BB/TB	-0.39	-1.46	-0.14	:	-0.35	-1.04	-1.37
Naik Berat Badan	O	O	O	:	-	-	-
PMT Diterima(kg)	-	0.84	-	:	-	-	-
Jml Vit A	-	-	-	:	-	-	-
KPSP	-	-	-	:	-	-	-
KIA	-	-	-	:	-	-	-

3.3 Seleksi Dan Integrasi Data

Tahap ini berjalan setelah data yang dikumpulkan kemudian dipilih atribut-atribut yang akan digunakan. Langkah ini bertujuan untuk meningkatkan akurasi dan efisiensi dari algoritma yang diterapkan. Proses seleksi dan integrasi pada tahap pertama melibatkan penyeleksian atribut-atribut yang dianggap tidak diperlukan. Atribut-atribut yang dihapus dari data stunting Kota Samarinda 2023 antara lain adalah BB Lahir, TB Lahir, Prov, Kab/Kota, Kec, Puskesmas, Desa/Kel, Posyandu, RT, RW, Usia Saat Ukur, PMT Diterima (kg), KPSP, dan KIA. Berikut adalah hasil integrasi tahap pertama setelah atribut yang tidak diperlukan dihapus: Nama, JK, Tanggal Pengukuran, Berat, Tinggi, LiLA, BB/U, ZS BB/U, TB/U, ZS TB/U, BB/TB, ZS BB/TB, Naik Berat Badan, dan Jml Vit A.

Tabel 3.2 Hasil Seleksi dan Integrasi Data

No	Nama	JK	Berat	Tinggi	LiLA	BB/UU	ZS BB/UU	TB/UU	ZS TB/UU	BB/TB	ZS BB/TB	NBB	Tanggal Pengukuran
1	Dimas Aditya	L	9.1	74.5	0	BB Normal	-0.39	Normal	-0.21	Gizi Baik	-0.39	O	2023-01-02
2	Siti Aisyah	P	12	94	0	Kurang	-2.25	Pendek	-2.09	Gizi Baik	-1.46	O	2023-01-02
3	M Al Fatih	L	8.1	69	0	BB Normal	-0.53	Normal	-0.65	Gizi Baik	-0.14	O	2023-01-02
:	:	:	:	:	:	:	:	:	:	:	:	:	:
150473	Afizah Khairina	P	2.5	45	-	Kurang	-2.03	Pendek	-2.63	Gizi Baik	-0.35	-	2023-12-19
150474	M Arsyah Kholif	P	3	49	-	BB Normal		Normal		Gizi Baik	-1.04	-	2023-12-19
150475	Muhammad Iqbal	L	2.9	49	-	Kurang		Pendek		Gizi Baik	-1.37	-	2023-12-19

Tabel 3.1 menunjukkan hasil seleksi dan integrasi data setelah menghapus atribut yang tidak relevan. Data yang dipilih mencakup atribut penting seperti nama, jenis kelamin, berat badan, tinggi badan, lingkaran lengan atas, dan berbagai *z-score* terkait status gizi anak. Proses ini bertujuan untuk meningkatkan akurasi dan efisiensi model klasifikasi dengan fokus pada atribut yang signifikan dalam analisis stunting di Kota Samarinda.

3.3.1 Data Cleaning

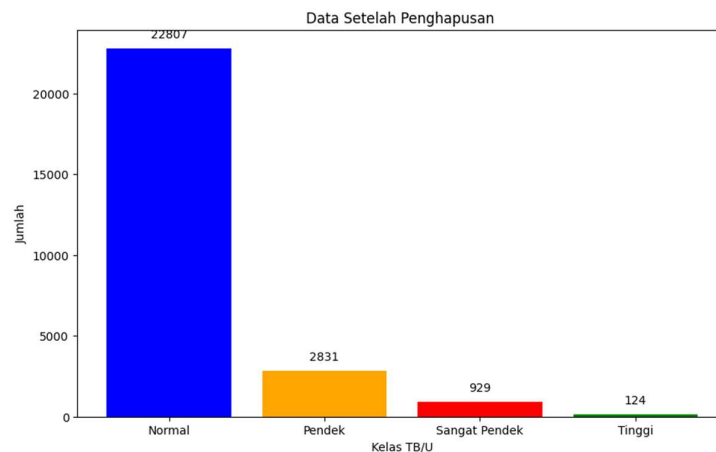
Tahap selanjutnya akan melibatkan pembersihan data untuk memastikan tidak adanya anomali atau data yang hilang (missing data), serta data yang terduplikasi, sehingga proses analisis dapat berjalan lebih efisien dan hasilnya lebih akurat. Dengan data yang sudah diseleksi dan diintegrasikan dengan baik, langkah ini akan membantu dalam mendapatkan hasil analisis yang lebih baik. Hasil data Cleaning terlihat pada table 3.2.

Tabel 3.3 Data Setelah Cleaning

No	Nama	JK	Berat	Tinggi	LiLA	BB/UU	ZS BB/UU	TB/UU	ZS TB/UU	BB/TB	ZS BB/TB	NBB	Tanggal Pengukuran
1	A	L	18.6	104.5	15	Risiko	1.19	Normal	0.41	Risiko Gizi	1.41	N	2023-12-16

	Alvin					Lebih				Lebih			
2	A Fadlan	L	11.6	83	0	BB Normal	-0.08	Normal	-0.97	Gizi Baik	0.59	T	2023-07-10
:	:	:	:	:	:	:	:	:	:	:	:	:	:
26690	Zubair	L	17.3	107	0	BB Normal	-0.14	Normal	-0.13	Gizi Baik	-0.13	O	2023-02-20
26691	Zulkifli Abdi	L	15.6	102	0	BB Normal	-0.59	Normal	-0.69	Gizi Baik	-0.25	N	2023-12-03

Pada Tabel 3.2, pembersihan data yang dilakukan pada dataset stunting Kota Samarinda mencakup penghapusan data yang memiliki nilai #N/A (missing data) atau data yang terduplikasi. Dalam proses ini, total data yang dihapus karena terduplikasi adalah 116.273 data, dan total data yang dihapus karena memiliki nilai #N/A adalah 7.510 data. Sehingga, setelah proses pembersihan data, jumlah data yang tersisa adalah 26.691.



Gambar 3.1 Dataset Setelah Tahap *Cleaning*

Pada Gambar 3.2 menunjukkan distribusi data stunting Kota Samarinda setelah pembersihan data sesuai Tabel 3.2. Dari total 26.691 record yang tersisa, sebagian besar anak berada dalam kategori tinggi badan normal (22.807), sementara sisanya tersebar dalam kategori pendek (2.831), sangat pendek (929), dan tinggi (124). Data ini mencerminkan mayoritas anak memiliki tinggi badan normal, namun masih ada yang mengalami stunting.

3.3.2 Data Transformation

Dalam tahap ini, Transformasi data merupakan langkah penting dalam proses analisis data, di mana atribut kategorikal diubah menjadi numerik untuk memungkinkan pemrosesan yang lebih efektif oleh algoritma machine learning. Pada langkah ini, kita menggunakan teknik seperti Label Encoding dan Ordinal Encoding untuk mengubah nilai kategorikal seperti 'L' dan 'P' dalam kolom 'JK' menjadi nilai numerik. Proses ini tidak hanya meningkatkan efisiensi komputasi tetapi juga memastikan bahwa data dapat diolah oleh model machine learning yang umumnya membutuhkan input numerik. Atribut yang akan ditransformasi mencakup JK, BB/UU, BB/TB, Naik Berat Badan dan TB/UU.

Tabel 3.4 Data Sebelum di *Transformation*

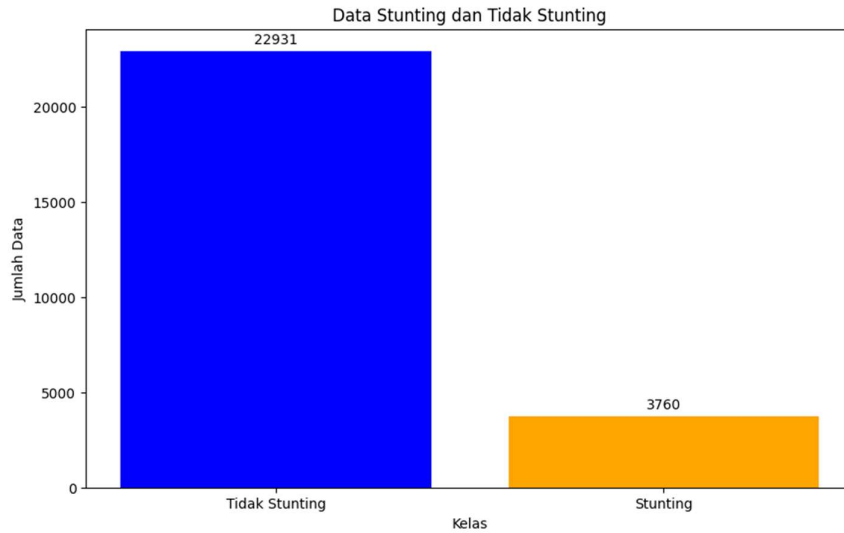
No	JK	BB/UU	BB/TB	Naik Berat Badan	TB/UU
1	L	Risiko Lebih	Risiko Gizi Lebih	N	Normal
2	L	BB Normal	Gizi Baik	T	Normal
3	L	BB Normal	Gizi Baik	O	Pendek
:	:	:	:	:	:
26689	P	Risiko Lebih	Risiko Gizi Lebih	0	Normal
26690	L	BB Normal	Gizi Baik	O	Normal
26691	L	BB Normal	Gizi Baik	N	Normal

Pada Tabel 3.3, saya memilih kolom JK, BB/UU, BB/TB, dan NBB. Untuk mentransformasi datanya menjadi numerik.

Tabel 3.5 Data Setelah di *Transformation*

No	JK	BB/UU	BB/TB	Naik Berat Badan	Kelas
1	0	2	5	1	Tidak Stunting
2	0	0	0	3	Tidak Stunting
3	0	0	0	2	Stunting
:	:	:	:	:	:
26689	1	2	5	2	Tidak Stunting
26690	0	0	0	2	Tidak Stunting
26691	0	0	0	1	Tidak Stunting

Pada tabel di atas, atribut 'TB/UU' tidak ada karena atribut tersebut ingin saya gunakan sebagai kelas. Atribut 'TB/UU' akan dijadikan kelas untuk menentukan apakah seorang anak mengalami stunting atau tidak, dengan kategori 'Normal' dan 'Tinggi' diklasifikasikan sebagai 'Tidak Stunting', serta kategori 'Pendek' dan 'Sangat Pendek' diklasifikasikan sebagai 'Stunting'.



Gambar 3.2 Data Setelah diubah kategorinya

Pada gambar 3.2 menunjukkan bahwa dari total data, 22.931 anak dikategorikan sebagai "Tidak Stunting" (dengan tinggi badan per-usia normal atau tinggi), sedangkan 3.760 anak dikategorikan sebagai "Stunting" (dengan tinggi badan per-usia pendek atau sangat pendek).

3.4 Hasil Pemodelan Setiap Algoritma

Pada bagian ini, akan dijelaskan tahapan-tahapan dalam melakukan klasifikasi data stunting di Kota Samarinda menggunakan metode *Random Forest*. Proses pemodelan dimulai dengan penerapan metode *Random Forest* dasar untuk memperoleh gambaran awal dari hasil klasifikasi.

Tabel 3.6 Hasil Seleksi Fitur RFE

No	1	2	3	4	5	6	7	8	9	10
Fitur	ZS TB/U	ZS BB/U	BB/U	Berat	ZS BB/TB	BB/TB	Tinggi	LiLA	Naik Berat Badan	JK
Nilai	0.647	0.157	0.094	0.031	0.030	0.196	0.155	0.001	0.0006	0.0005

Berdasarkan tabel di atas, dapat dilihat bahwa fitur-fitur yang memiliki pengaruh terbesar terhadap klasifikasi stunting adalah ZS TB/U, ZS BB/U, dan BB/U. ZS TB/U memiliki nilai kepentingan tertinggi, yaitu 0.647, yang berarti fitur ini memberikan kontribusi paling besar dalam model klasifikasi. Terakhir, dilakukan optimasi lebih lanjut pada model dengan menggabungkan seleksi fitur RFE dan optimasi menggunakan *Genetic Algorithm (GA)* untuk mencapai hasil klasifikasi yang lebih optimal.

3.4.1 *Random Forest*

Tahap awal, model *Random Forest* diterapkan untuk mengklasifikasikan data stunting di Kota Samarinda. Kemudian dilakukan pengujian menggunakan teknik validasi silang *10-fold cross-validation* untuk memastikan model diuji secara menyeluruh dan memberikan gambaran akurasi pada data yang berbeda. Setiap lipatan (*fold*) berfungsi sebagai data uji satu

kali. Hasil akurasi dari setiap fold dicatat untuk mengevaluasi kinerja model dalam mengklasifikasikan data stunting.

Tabel 3.7 Hasil Akurasi *Random Forest*

Fold	Akurasi
1	70.97%
2	86.08%
3	99.69%
4	89.48%
5	96.84%
6	91.66%
7	94.29%
8	96.66%
9	99.29%
10	94.23%

Dari hasil pengujian tersebut, terlihat bahwa akurasi model bervariasi pada setiap lipatan, dengan akurasi terendah sebesar 70.96% dan akurasi tertinggi mencapai 99.68%. Untuk lebih jelasnya di jelaskan dalam table berikut.

Tabel 3.8 Penjelasan *Random Forest K-Fold = 10*

Fold	Keterangan
1	Akurasi sebesar 70.95%, menunjukkan performa yang cukup baik meskipun merupakan akurasi terendah di antara semua fold.
2	Akurasi meningkat menjadi 86.07%, menunjukkan performa yang lebih baik.
3	Akurasi mencapai puncaknya di 99.67%, menunjukkan bahwa model hampir sempurna dalam klasifikasi pada data ini.
4	Akurasi sedikit menurun menjadi 89.47%, tetapi masih menunjukkan performa yang kuat.
5	Akurasi sangat tinggi sebesar 96.86%, mendekati performa optimal.
6	Akurasi sebesar 91.66%, tetap menunjukkan performa yang sangat baik.

7	Akurasi sebesar 94.27%, kembali menunjukkan performa yang kuat.
8	Akurasi sangat tinggi sebesar 96.67%, mendekati performa optimal.
9	Akurasi hampir sempurna sebesar 99.28%, sangat mendekati performa terbaik.
10	Akurasi sebesar 94.28%, menunjukkan hasil yang konsisten dan kuat.

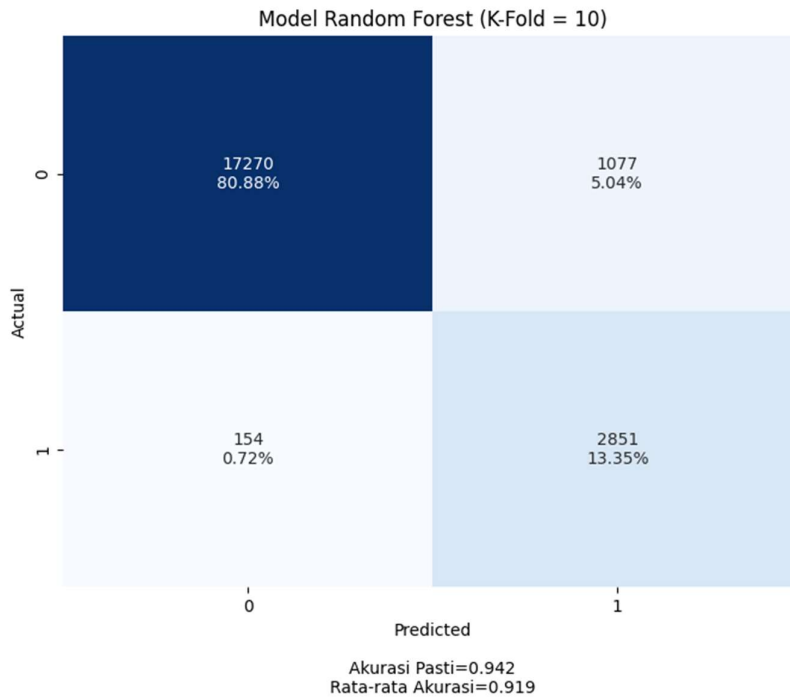
$$\text{Rata - Rata Accuracy} = \frac{919.19}{10} = 91.919\%$$

Rata-rata akurasi dari seluruh lipatan menunjukkan bahwa model Random Forest memiliki kinerja yang baik dalam mengklasifikasikan data stunting. Akurasi rata-rata model Random Forest dalam mengklasifikasikan data stunting Kota Samarinda Tahun 2023 adalah sebesar 91.91%.

Tabel 3.9 Confusion Matrix Random Forest

	<i>Predicted Positive (0)</i>	<i>Predicted Negative (1)</i>
<i>Actual Positive (0)</i>	17270	1077
<i>Actual Negative (1)</i>	154	2851

Tabel 3.7 menunjukkan Confusion Matrix untuk model Random Forest. Matriks ini membantu mengukur performa model klasifikasi dengan menunjukkan jumlah prediksi yang benar dan salah untuk masing-masing kelas. Dari tabel tersebut, terlihat bahwa ada 17,270 kasus True Positive (TP) di mana model memprediksi "Positive" (tidak stunting) dengan benar, 1,077 kasus False Negative (FN) di mana model salah memprediksi "Negative" (stunting), 154 kasus False Positive (FP) di mana model salah memprediksi "Positive" (tidak stunting), dan 2,851 kasus True Negative (TN) di mana model memprediksi "Negative" (stunting) dengan benar. Matriks ini digunakan untuk menghitung akurasi model, yang menunjukkan bahwa model memiliki tingkat keandalan yang tinggi dalam mengklasifikasikan data stunting.



Gambar 3.3 Confusion Matrix Random Forest

$$Accuracy = \frac{17270 + 2851}{17270 + 2851 + 154 + 1077} \times 100\% = 94.23\%$$

Dengan demikian, berdasarkan perhitungan manual diatas dan juga dengan perhitungan menggunakan kode *Python* akurasinya adalah 94.23%. Angka ini menunjukkan bahwa model *Random Forest* memiliki tingkat keandalan yang tinggi dalam mengklasifikasikan data stunting. Akurasi sebesar 94.23% mengindikasikan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar, dengan kesalahan yang relatif kecil.

3.4.2 Random Forest Dengan Seleksi Fitur *Recursive Feature Elimination (RFE)*

Setelah mendapatkan hasil awal dari penerapan metode *Random Forest*, langkah selanjutnya adalah meningkatkan performa model dengan teknik seleksi fitur. *Recursive Feature Elimination (RFE)* digunakan untuk memilih fitur-fitur yang paling relevan dan berpengaruh dalam klasifikasi data stunting. *RFE* bekerja dengan menghapus fitur secara berulang-ulang dan membangun model pada fitur yang tersisa untuk mengidentifikasi kombinasi fitur terbaik. Dengan menerapkan *RFE*, diharapkan model *Random Forest* akan lebih efisien dan akurat dalam melakukan klasifikasi. Proses ini melibatkan evaluasi kinerja model setelah seleksi fitur dan membandingkannya dengan model dasar untuk menentukan peningkatan yang diperoleh.

Tabel 3.10 Hasil Akurasi Random Forest + RFE

Fold	Akurasi
1	81.72%

2	84.93%
3	99.51%
4	86.45%
5	98.53%
6	93.52%
7	100%
8	95.00%
9	99.66%
10	97.10%

Dari hasil pengujian tersebut, terlihat bahwa akurasi model bervariasi pada setiap lipatan, dengan akurasi terendah sebesar 81.72% dan akurasi tertinggi mencapai 100%. Untuk lebih jelasnya di jelaskan dalam table berikut.

Tabel 3.11 Penjelasan *RF + RFE K-Fold = 10*

Fold	Keterangan
1	Akurasi sebesar 81.72%, menunjukkan performa yang cukup baik pada subset data ini.
2	Akurasi meningkat menjadi 84.93%, menunjukkan perbaikan kinerja model.
3	Akurasi mencapai 99.51%, mendekati performa optimal dan menunjukkan kemampuan model yang sangat baik dalam mengklasifikasikan data pada subset ini.
4	Akurasi sebesar 86.45%, mempertahankan performa yang kuat.
5	Akurasi sangat tinggi sebesar 98.53%, mendekati akurasi optimal.
6	Akurasi sebesar 93.52%, menunjukkan kinerja yang sangat baik.
7	Akurasi sempurna sebesar 100%, menunjukkan bahwa model mampu mengklasifikasikan semua data dengan benar pada subset ini.
8	Akurasi sebesar 95.00%, kembali menunjukkan performa yang kuat.
9	Akurasi sangat tinggi sebesar 99.66%, sangat

	mendekati performa terbaik.
10	Akurasi sebesar 97.10%, menunjukkan hasil yang konsisten dan kuat.

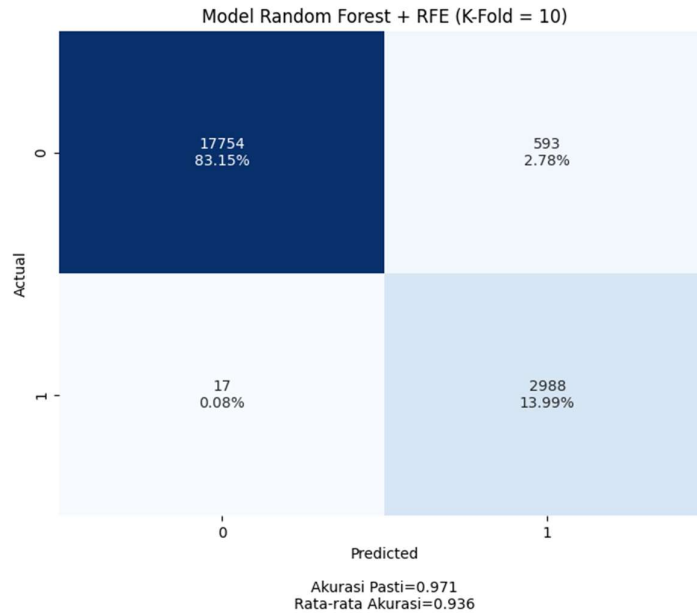
$$\text{Rata - Rata Accuracy} = \frac{936.42}{10} = 93.64\%$$

Rata-rata akurasi dari seluruh lipatan menunjukkan bahwa model Random Forest dengan seleksi fitur Recursive Feature Elimination memiliki kinerja yang baik dalam mengklasifikasikan data stunting. Akurasi rata-rata model *RF + RFE* dalam mengklasifikasikan data stunting Kota Samarinda Tahun 2023 adalah sebesar 93.64%.

Tabel 3.12 Confusion Matrix *RF + RFE*

	<i>Predicted Positive (0)</i>	<i>Predicted Negative (1)</i>
<i>Actual Positive (0)</i>	17754	593
<i>Actual Negative (1)</i>	17	2988

Tabel 3.10 menunjukkan Confusion Matrix untuk model Random Forest (RF) yang telah menggunakan Recursive Feature Elimination (RFE). Matriks ini menunjukkan performa model dengan rincian sebagai berikut: 17,754 kasus True Positive (TP) di mana model memprediksi "Positive" (tidak stunting) dengan benar, 593 kasus False Negative (FN) di mana model salah memprediksi "Negative" (stunting), 17 kasus False Positive (FP) di mana model salah memprediksi "Positive" (tidak stunting), dan 2,988 kasus True Negative (TN) di mana model memprediksi "Negative" (stunting) dengan benar. Confusion Matrix ini menunjukkan peningkatan akurasi dan keandalan model dalam mengklasifikasikan data stunting setelah penerapan seleksi fitur RFE.



Gambar 3.4 Confusion Matrix RF + RFE

$$Accuracy = \frac{17754 + 2988}{17754 + 2988 + 17 + 593} \times 100\% = 97.10\%$$

Dengan demikian, berdasarkan perhitungan manual diatas dan juga dengan perhitungan menggunakan kode *Python* akurasinya adalah 97.10%. Angka ini menunjukkan bahwa model *Random Forest* dengan seleksi fitur *Recursive Feature Elimination* memiliki tingkat keandalan yang tinggi dalam mengklasifikasikan data stunting. Akurasi sebesar 97.10% mengindikasikan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar, dengan kesalahan yang relatif kecil.

3.4.3 *Random Forest* Dengan Seleksi Fitur *RFE* Dan Optimasi *GA* (*Genetic Algorithm*)

Pada tahap ini, model *Random Forest* dikombinasikan dengan teknik seleksi fitur *Recursive Feature Elimination (RFE)* dan optimasi *Genetic Algorithm (GA)* untuk meningkatkan akurasi klasifikasi data stunting di Kota Samarinda. Proses ini melibatkan dua langkah utama: pertama, seleksi fitur menggunakan *RFE* untuk memilih fitur-fitur yang paling relevan, dan kedua, optimasi *hyperparameter* model menggunakan *Genetic Algorithm* untuk menemukan kombinasi parameter yang optimal. Kombinasi dari kedua teknik ini bertujuan untuk meningkatkan kinerja model secara keseluruhan dengan mengurangi kompleksitas data dan menemukan konfigurasi terbaik untuk model *Random Forest*.

Tabel 3.13 Hasil Akurasi *Random Forest* + *RFE* + *GA*

Fold	Akurasi
1	99.62%
2	99.73%

3	89.31%
4	95.50%
5	99.73%
6	99.82%
7	99.50%
8	100%
9	99.99%
10	99.70%

Dari hasil pengujian tersebut, terlihat bahwa akurasi model bervariasi pada setiap lipatan, dengan akurasi terendah sebesar 89.31% dan akurasi tertinggi mencapai 100%. Untuk lebih jelasnya di jelaskan dalam table berikut.

Tabel 3.14 Penjelasan *RF + RFE + GA K-Fold = 10*

Fold	Keterangan
1	Akurasi sangat tinggi sebesar 99.62%, menunjukkan model yang sangat baik.
2	Akurasi sedikit meningkat menjadi 99.73%, mempertahankan performa yang sangat kuat.
3	Akurasi lebih rendah di 89.31%, tetapi masih dalam kisaran yang baik.
4	Akurasi sebesar 95.50%, menunjukkan peningkatan performa yang signifikan.
5	Akurasi kembali sangat tinggi sebesar 99.73%.
6	Akurasi mencapai 99.82%, mendekati performa sempurna.
7	Akurasi sebesar 99.50%, tetap menunjukkan hasil yang sangat baik.
8	Akurasi sempurna sebesar 100%, menunjukkan bahwa model mampu mengklasifikasikan semua data dengan benar pada subset ini.
9	Akurasi hampir sempurna sebesar 99.99%.
10	Akurasi sangat tinggi sebesar 99.70%, menunjukkan hasil yang konsisten.

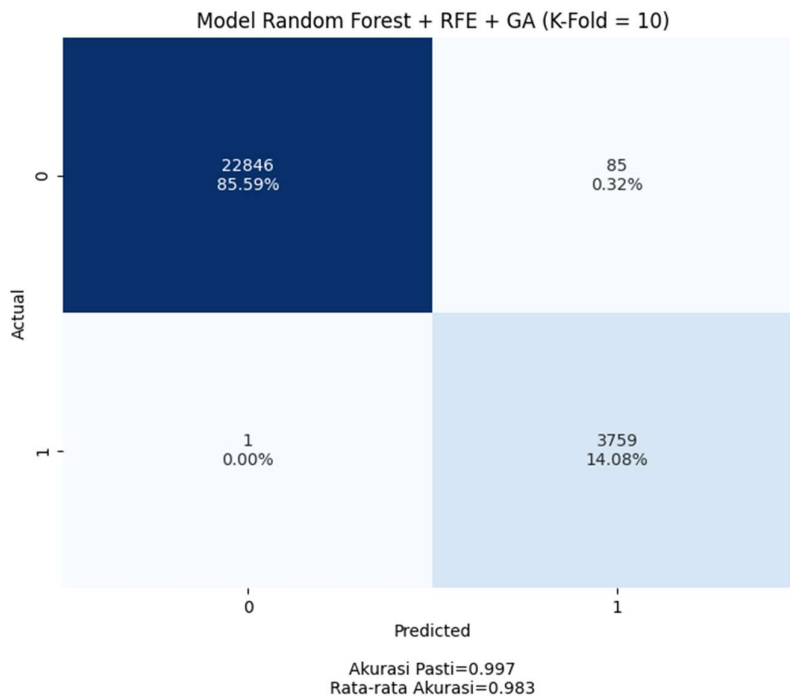
$$\text{Rata - Rata Accuracy} = \frac{983.90}{10} = 98.39\%$$

Rata-rata akurasi dari seluruh lipatan adalah sekitar 98.39%, yang menunjukkan bahwa kombinasi model *Random Forest* dengan seleksi fitur *RFE* dan optimasi *GA* memiliki kinerja yang sangat baik dalam mengklasifikasikan data stunting. Hasil ini mencerminkan peningkatan yang signifikan dan keandalan model dalam berbagai subset data, dengan akurasi yang sangat tinggi di setiap fold. Performa model yang kuat ini memberikan kepercayaan yang tinggi terhadap kemampuannya dalam aplikasi nyata untuk mengidentifikasi kasus stunting di Kota Samarinda.

Tabel 3.15 Confusion Matrix RF + RFE + GA

	Predicted Positive (0)	Predicted Negative (1)
Actual Positive (0)	22846	85
Actual Negative (1)	1	3759

Tabel 3.13 menunjukkan Confusion Matrix untuk model Random Forest (RF) yang telah menggunakan Recursive Feature Elimination (RFE) dan Genetic Algorithm (GA). Matriks ini menunjukkan hasil sebagai berikut: 22,846 kasus True Positive (TP) di mana model memprediksi "Positive" (tidak stunting) dengan benar, 85 kasus False Negative (FN) di mana model salah memprediksi "Negative" (stunting), 1 kasus False Positive (FP) di mana model salah memprediksi "Positive" (tidak stunting), dan 3,759 kasus True Negative (TN) di mana model memprediksi "Negative" (stunting) dengan benar. Confusion Matrix ini menunjukkan performa tinggi dan peningkatan signifikan dalam akurasi dan keandalan model setelah penerapan RFE dan optimasi GA.



Gambar 3.5 Confusion Matrix RF + RFE + GA

$$Accuracy = \frac{22846 + 3759}{22846 + 3759 + 1 + 85} \times 100\% = 99.70\%$$

Dengan demikian, berdasarkan perhitungan manual diatas dan juga dengan perhitungan menggunakan kode *Python* akurasi adalah 98.39%. Angka ini menunjukkan bahwa model *Random Forest* dengan seleksi fitur *RFE* dan optimasi *GA* memiliki tingkat keandalan yang tinggi dalam mengklasifikasikan data stunting. Akurasi sebesar 99.70% mengindikasikan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar, dengan kesalahan yang relatif kecil.

3.4.4 Perbandingan Hasil Akurasi

Pada bagian ini, akan dilakukan perbandingan hasil akurasi dari tiga metode yang telah diterapkan untuk mengklasifikasikan data stunting di Kota Samarinda, yaitu *Random Forest (RF)*, *Random Forest* dengan Seleksi Fitur *RFE (RF+RFE)*, dan *Random Forest* dengan Seleksi Fitur *RFE* serta *Optimasi GA (RF+RFE+GA)*. Hasil akurasi dari setiap metode akan ditampilkan dalam bentuk tabel untuk memudahkan analisis dan evaluasi. Dengan membandingkan ketiga metode ini, dapat diketahui seberapa besar peningkatan performa yang diperoleh dari setiap tahapan seleksi fitur dan optimasi yang dilakukan. Tabel berikut menyajikan ringkasan hasil akurasi dari masing-masing metode:

Tabel 3.16 Perbandingan Hasil Akurasi Sesudah Penggunaan *RFE & GA*

Fold	<i>RF</i>	<i>RF + RFE</i>	<i>RF + RFE + GA</i>	Perubahan <i>RF</i> ke <i>RF-RFE</i>	Perubahan <i>RF</i> ke <i>RF-RFE-GA</i>	Perubahan <i>RF-RFE</i> ke <i>RF-RFE-GA</i>
1	70.97%	81.72%	99.62%	10.75%	28.65%	17.90%
2	86.08%	84.93%	99.73%	-1.15%	13.65%	14.80%
3	99.69%	99.51%	89.31%	-0.18%	-10.38%	-10.20%
4	89.48%	86.45%	95.50%	-3.03%	6.02%	9.05%
5	96.84%	98.53%	99.73%	1.69%	2.89%	1.20%
6	91.66%	93.52%	99.82%	1.86%	8.16%	6.30%
7	94.29%	100%	99.50%	5.71%	5.21%	-0.50%
8	96.66%	95.00%	100%	-1.66%	3.34%	5%
9	99.29%	99.66%	99.99%	0.37%	0.70%	0.33%
10	94.23%	97.10%	99.70%	2.87%	5.47%	2.60%

Tabel 3.15 menunjukkan perbandingan hasil klasifikasi menggunakan *Random Forest (RF)*, *Random Forest* dengan Seleksi Fitur (*RF + RFE*), dan *Random Forest* dengan Seleksi Fitur dan Optimasi (*RF + RFE + GA*) pada 10 *fold* validasi, dengan sebagian besar *fold* menunjukkan peningkatan akurasi setelah seleksi fitur dan optimasi, meskipun terdapat beberapa penurunan akurasi pada *fold* tertentu yang dapat disebabkan oleh fitur yang tidak relevan, *overfitting*, atau variasi acak dalam data.

Tabel 3.17 Perbandingan Hasil Rata-Rata Akurasi Sesudah Penggunaan *RFE* & *GA*

Average Accuracy	<i>RF</i>	<i>RF</i> + <i>RFE</i>	<i>RF</i> + <i>RFE</i> + <i>GA</i>	Perubahan <i>RF</i> ke <i>RF-RFE</i>	Perubahan <i>RF</i> ke <i>RF-RFE-GA</i>	Perubahan <i>RF-RFE</i> ke <i>RF-RFE-GA</i>
	91.91%	93.64%	98.39%	1.73%	6.48%	4.75%

Tabel 3.15 dan 3.16 menunjukkan perbandingan hasil klasifikasi data stunting di Kota Samarinda menggunakan tiga metode berbeda: *Random Forest (RF)*, *Random Forest* dengan Seleksi Fitur *RFE (RF + RFE)*, dan *Random Forest* dengan Seleksi Fitur *RFE* serta Optimasi *GA (RF + RFE + GA)*. Tabel ini memberikan gambaran jelas tentang bagaimana setiap metode mempengaruhi akurasi klasifikasi, Berikut adalah penjelasan tiap hasilnya.

- Pada metode *RF* dasar, rata-rata akurasi yang diperoleh adalah 91.91%, sedangkan akurasi pasti mencapai 94.23%. Dengan penambahan seleksi fitur menggunakan *RFE*, rata-rata akurasi meningkat menjadi 93.64%, yang menunjukkan peningkatan sebesar 1.73%. Akurasi pasti juga meningkat menjadi 97.10%, menunjukkan peningkatan sebesar 2.87%.
- Selanjutnya, dengan menggabungkan *RFE* dan optimasi *GA*, rata-rata akurasi meningkat signifikan menjadi 98.39%, menunjukkan peningkatan sebesar 4.75% dibandingkan metode *RF* dasar. Akurasi pasti mencapai nilai tertinggi sebesar 99.70%, dengan peningkatan sebesar 2.60% dibandingkan metode *RF + RFE*.

3.5 Pembahasan

Penelitian ini telah melalui serangkaian tahapan mulai dari pengumpulan data, pembersihan data, transformasi data, hingga pemodelan menggunakan berbagai teknik machine learning untuk mengklasifikasikan data stunting di Kota Samarinda. Hasil dari setiap tahapan ini menunjukkan peningkatan akurasi yang signifikan seiring dengan penerapan teknik seleksi fitur dan optimasi. Data awal terdiri dari 27 atribut yang diperoleh dari Dinas Kesehatan Kota Samarinda. Melalui proses seleksi fitur, atribut yang kurang relevan dihapus, menyisakan 12 atribut sebagai fitur dan 1 atribut sebagai target dalam klasifikasi stunting. Proses seleksi ini bertujuan untuk meningkatkan efisiensi dan akurasi model dengan fokus pada atribut yang signifikan. Data yang diperoleh melalui proses pembersihan untuk menghapus data yang tidak lengkap dan terduplikasi. Selanjutnya, transformasi data dilakukan untuk mengubah atribut kategorikal menjadi numerik, sehingga dapat diolah oleh algoritma machine learning.

Model awal menggunakan *Random Forest* menunjukkan akurasi yang cukup baik dengan rata-rata akurasi sebesar 91.91% dan akurasi pasti 94.23%. Ini menunjukkan bahwa model dasar sudah cukup efektif dalam mengklasifikasikan data stunting. Penggunaan *RFE* untuk seleksi fitur meningkatkan akurasi model. Rata-rata akurasi meningkat menjadi 93.64%, dengan akurasi pasti 97.10%. Ini menunjukkan bahwa *RFE* membantu dalam mengidentifikasi fitur yang paling relevan, sehingga meningkatkan performa model. Kombinasi *RFE* dan optimasi *GA* menghasilkan peningkatan akurasi yang signifikan. Rata-rata akurasi mencapai 98.39% dengan akurasi pasti 99.70%. Ini menunjukkan bahwa *GA* efektif dalam menemukan konfigurasi *hyperparameter* yang optimal untuk model *Random Forest*.

Perbandingan dengan penelitian terdahulu menunjukkan peningkatan yang signifikan dalam performa model. Sebagai contoh, penelitian sebelumnya oleh (Obvious et al. 2023) menggunakan algoritma Support Vector Classification dan XGBoost dengan akurasi masing-masing sekitar 64% dan 63%. Selain itu, penelitian (Estiyak et al. 2023) menunjukkan bahwa algoritma Random Forest, Naive Bayes, dan Logistic Regression mengalami penurunan performa saat dihadapkan pada data berdimensi tinggi, dengan akurasi masing-masing 60%, 58%, dan 58%. Dalam konteks data berdimensi tinggi, penelitian ini menunjukkan bahwa penggunaan teknik seleksi fitur RFE dan optimasi GA mampu meningkatkan akurasi model secara signifikan dibandingkan dengan pendekatan tradisional yang digunakan dalam penelitian terdahulu. (Leykun et al. (2024) menunjukkan bahwa penggunaan RFE dapat meningkatkan akurasi dari 72.41% menjadi 80.1% untuk metode Random Forest, sementara penelitian ini lebih jauh menunjukkan bahwa dengan tambahan optimasi GA, akurasi dapat meningkat hingga 99.70%. Dan untuk menjawab rumusan masalah pada bab1 berikut penjelasannya:

- a) Penelitian ini bertujuan untuk mengidentifikasi fitur-fitur yang paling relevan dan signifikan dalam klasifikasi data stunting. Proses seleksi fitur menggunakan *RFE* diharapkan dapat mengeliminasi fitur yang kurang relevan, sehingga model dapat bekerja lebih efisien dan akurat. Dalam konteks ini, fitur-fitur yang memiliki pengaruh terbesar terhadap klasifikasi stunting adalah ZS TB/U, ZS BB/U, dan BB/U. ZS TB/U memiliki nilai kepentingan tertinggi, yaitu 0.647, yang berarti fitur ini memberikan kontribusi paling besar dalam model klasifikasi. Berikut ini adalah tabel yang memperlihatkan perbandingan fitur yang digunakan dalam penelitian ini dengan penelitian lain:

Tabel 3.18 Perbandingan *RFE* Dengan Penelitian lain

Penelitian	Metode	Data	Fitur	Akurasi
Penelitian Ini	<i>RF + RFE</i>	Stunting	ZS TB/U, ZS BB/U, dan BB/U.	94.23% - 97.10% (+ 2.87%)
(Yoga et al., 2021)	<i>SVM + Chi-Square</i>	Stunting	ZS TB/U, BB/U, ZS BB/U, Tinggi, LiLA, dan Berat.	96.50% - 96.60% (+ 0.10%)
(Kemal et al., 2023)	<i>K-NN + Relief-F</i>	Stunting	Umur, Berat, Tinggi, BB/U, ZS BB/U, BB/TB, ZS BB/TB, dan ZS TB/U.	91.90% - 92.20% (+ 0.30%)
(Syahrani dan Dwi, 2022)	<i>K-NN + Backward Elimination</i>	Stunting	Usia Saat Ukur, Berat, Tinggi, ZS BB/U dan ZS TB/U.	95.17% - 90.23% (- 4.94%)

- b) Setelah seleksi fitur dilakukan menggunakan *RFE*, langkah berikutnya adalah optimasi model menggunakan *GA* untuk meningkatkan akurasi prediksi. Hasil penelitian menunjukkan bahwa seleksi fitur menggunakan metode *RFE* terbukti efektif dalam meningkatkan kinerja model *Random Forest*. Dengan mengidentifikasi dan mengeliminasi fitur yang kurang relevan, rata-rata akurasi model meningkat dari 91.91% menjadi 93.64%. Akurasi pasti juga mengalami

peningkatan dari 94.23% menjadi 97.10%. Optimasi lebih lanjut dengan menggunakan *Genetic Algorithm (GA)* setelah seleksi fitur *RFE* memberikan peningkatan akurasi yang signifikan. Rata-rata akurasi model mencapai 98.39%, dan akurasi pasti mencapai 99.70%. Kombinasi *RFE* dan *GA* memungkinkan model *Random Forest* untuk mencapai performa maksimal dalam mengklasifikasikan data stunting, menunjukkan bahwa teknik seleksi fitur dan optimasi hyperparameter yang digunakan secara bersamaan dapat secara efektif mengatasi kompleksitas data dan meningkatkan keandalan prediksi stunting. Berikut ini adalah tabel yang memperlihatkan perbandingan fitur yang digunakan dalam penelitian ini dengan penelitian lain:

Tabel 3.19 Perbandingan *GA* Dengan Penelitian lain

Penelitian	Metode	Akurasi
Penelitian Ini	<i>RF + RFE + GA</i>	99.70%
(Monire et al., 2021)	<i>PSO + RNN</i>	96.08%
	<i>RF + SMOTE</i>	98.31%
	<i>PSO + KNN</i>	98.9%
	<i>EGA + PSO</i>	98.97%
	<i>PSO + RF</i>	99.76%
	<i>GA + RF</i>	99.99%