

**IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK KLASIFIKASI  
RISIKO SERANGAN JANTUNG**

**SKRIPSI**

**Oleh:  
NADIA MU'IZZATUL ILMA  
NIM. 200605110075**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2024**

**IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK  
KLASIFIKASI RISIKO SERANGAN JANTUNG**

**SKRIPSI**

Diajukan kepada:  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:  
**NADIA MU'IZZATUL ILMA**  
**NIM. 200605110075**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2024**

**HALAMAN PERSETUJUAN**

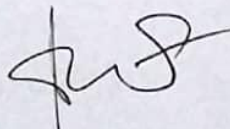
**IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK  
KLASIFIKASI RISIKO SERANGAN JANTUNG**

**SKRIPSI**

**Oleh:**  
**NADIA MU'IZZATUL ILMA**  
**NIM. 200605110075**

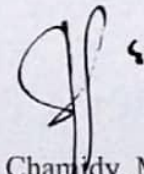
Telah Diperiksa dan Disetujui untuk Diuji  
Tanggal: 22 Mei 2024

Pembimbing I



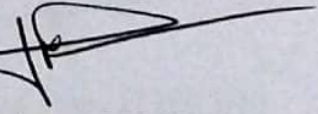
Prof. Dr. Suhartono, M.Kom  
NIP. 19680519 200312 1 001

Pembimbing II



Dr. Totok Chamidy, M.Kom  
NIP. 19691222 200604 1 001

Mengetahui,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Gegeri Maulana Malik Ibrahim Malang



Dr. Fachrul Kurniawan, M.MT, IPM  
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK  
KLASIFIKASI RISIKO SERANGAN JANTUNG

SKRIPSI

Oleh:  
**NADIA MU'IZZATUL ILMA**  
NIM. 200605110075

Telah dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk memperoleh Gelar Sarjana Komputer (S.Kom)  
Tanggal: 06 Juni 2024


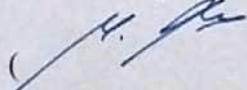
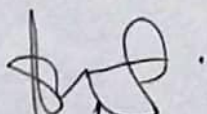
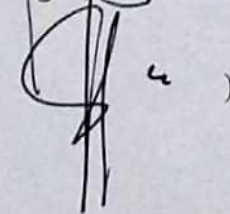
**Susunan Dewan Penguji**

Ketua Penguji : Dr. Muhammad Faisal, M.T  
NIP. 19740510 200501 1 007

Anggota Penguji I : Dr. M. Ainul Yaqin, M.Kom  
NIP. 19761013 200604 1 004

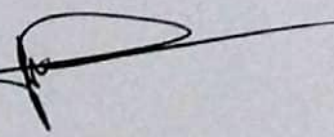
Anggota Penguji II : Prof. Dr. Suhartono, M.Kom  
NIP. 19680519 200312 1 001

Anggota Penguji III : Dr. Totok Chamidy, M.Kom  
NIP. 19691222 200604 1 001

(  )  
(  )  
(  )  
(  )

Mengetahui dan Mengesahkan,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
Dr. Fachrul Kurniawan, M.MT, IPM  
NIP. 19771020 200912 1 001



## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Nadia Mu'izzatul Ilma  
NIM : 200605110075  
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika  
Judul Skripsi : Implementasi Algoritma *Random Forest* untuk  
Klasifikasi Risiko Serangan Jantung

Menyatakan dengan aktualnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 06 Juni 2024  
Yang membuat pernyataan,



Nadia Mu'izzatul Ilma  
NIM. 200605110075

## HALAMAN MOTTO

“Apabila sesuatu yang kau senangi tidak terjadi maka senangilah apa yang terjadi”

(Ali bin Abi Thalib)

“Semua ada waktunya. Jangan membandingkan hidupmu dengan hidup orang lain. Tidak ada perbandingan antara matahari dan bulan, mereka bersinar saat waktunya tiba”

(B. J. Habibie)

*“You’re doing fine. Sometimes you’re doing better, sometimes you’re doing worse. But at the end, it’s you. I just want you to have no regrets, I want you to feel yourself grow, and I just want you, to love yourself”*

(Mark)

## **HALAMAN PERSEMBAHAN**

Alhamdulillah Wasyukurillah segala Puji Syukur atas kehadiran Allah SWT atas limpahan Rahmat, Taufiq, hidayah, dan inayah-Nya kepada penulis sehingga penulis dapat menyelesaikan tugas akhir ini. Dengan rasa hormat dan terimakasih, penulis mempersembahkan skripsi tugas akhir ini kepada:

1. Ibu Nur Afifah almarhumah. Terima kasih atas limpahan kasih sayang yang tak terhingga yang telah mengasuh, merawat, mendidik dan membesarkan penulis walaupun tidak bisa membersamai sampai di tahap ini, yang telah mengajari penulis artinya bertahan, nasihat untuk terus bisa berdiri di kaki sendiri, dan telah menjadi motivasi utama penulis untuk menyelesaikan studi sebagai bentuk mewujudkan impian ibu yang tak sempat tercapai.
2. Bapak Abdul Ghofar. Terima kasih atas limpahan kasih sayang yang tak terhingga yang telah mengasuh, merawat, mendidik dan membesarkan penulis hingga saat ini. Terima kasih untuk setiap doa, lelah, segala hal yang dilakukan dan diperjuangkan agar penulis bisa sampai di tahap ini, yang selalu mengajari penulis artinya perjuangan, bersyukur, dan bertahan. Terima kasih telah menjadi salah satu motivasi utama penulis menyelesaikan studi.
3. Bunda Istifadah. Terima kasih telah hadir dalam kehidupan penulis. Terima kasih bunda atas kasih sayang, perhatian, nasihat, dan semua yang telah diberikan kepada penulis. Terima kasih untuk mau turut mengasuh, mendidik, dan mengusahakan sampai penulis bisa di tahap ini.
4. Adikku Marsa Aulia Qonita, terima kasih atas semangat, masukan, sabar dalam menampung keluh kesah, menuruti keinginan penulis yang terkadang tiba-tiba, dan selalu menghibur penulis ketika lelah.
5. Dosen Pembimbing penulis yaitu Bapak Prof. Dr. Suhartono, M.Kom dan Bapak Dr. Totok Chamidy, M.Kom, yang sabar serta tulus dalam memberikan bimbingan dan arahan kepada penulis dalam penyelesaian tugas akhir.

## KATA PENGANTAR

*Assalamualaikum Wr. Wb*

Puji dan syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa, Allah subhanahu wa ta'ala yang telah memberikan Taufik dan Hidayah-Nya kepada penulis sehingga dapat menyelesaikan skripsi yang berjudul “Implementasi Algoritma *Random Forest* Untuk Klasifikasi Risiko Serangan Jantung”.

Para pihak yang telah memberikan dukungan kepada penulis. Dalam kesempatan kali ini penulis ingin mengucapkan banyak terimakasih kepada:

1. Prof. Dr. H. M. Zainuddin, MA, selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang beserta jajarannya.
2. Prof. Dr. Sri Harini, M.Si, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang beserta jajarannya.
3. Dr. Fachrul Kurniawan ST., M.MT, IPM selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Prof. Dr. Suhartono, M.Kom selaku Dosen Pembimbing I yang telah dengan sabar memberikan arahan baik dalam penulisan hingga program yang dibuat dalam menyelesaikan skripsi ini.
5. Dr. Totok Chamidy, M.Kom selaku Dosen Pembimbing II yang telah dengan sabar memberikan arahan baik dalam penulisan hingga program yang dibuat dalam menyelesaikan skripsi ini.
6. Dr. Cahyo Crysdiyan, M.Cs selaku Dosen Wali yang telah memberikan arahan dalam proses perkuliahan.
7. Nia Faricha, S.Si selaku Admin Program Studi Teknik Informatika yang dengan sabar membantu, memberikan arahan, informasi terkait perkuliahan.
8. Segenap dosen, laboran, dan jajaran staf pada Program Studi Teknik Informatika yang telah memberikan bimbingan dan bantuan selama studi.
9. Orangtua penulis almarhumah Ibu Nur Afifah, Bapak Abdul Ghofar, Bunda Istifadah yang selalu menjadi motivasi penulis untuk menyelesaikan studi, memberi dukungan dan perhatian, serta selalu memberikan yang terbaik untuk kelancaran putrinya dalam hal apapun.



10. Adik penulis Marsa Aulia Qonita yang selalu memberi semangat, menampung keluh kesah penulis, dan menghibur penulis saat merasa lelah.
11. Keluarga penulis terutama untuk Mba Nisa, Tante Via, *Mom* Elma, Ayah Jek, Tante Meme, dan yang lainnya. Terima kasih sudah selalu mendukung, menyemangati, memberi nasihat, dan segala apapun yang telah diberikan untuk penulis.
12. Sahabat penulis penghuni grup “*Gantadrie*” yaitu Kartika dan Rila, terima kasih atas segala bantuan, segala keluh kesah yang ditampung, dan semangat yang diberikan serta semua kenangan dari awal perkuliahan yang masih *online*, bertemu hanya lewat media sosial sampai akhirnya bisa bertemu secara langsung dan saling membantu dalam perkuliahan, sampai terselesaikannya skripsi ini.
13. Teman-teman penulis Dinda, Vera, Alfina, Rizka, Ahida, anggota divisi seni dan olahraga Himatif 2022, dan yang tidak bisa disebutkan satu persatu, terima kasih sudah membantu, mendukung serta menghibur penulis di masa-masa hilang harapan. Selalu sigap disaat penulis membutuhkan pertolongan, dan banyak hal lain.
14. Keluarga besar Saudara Teknik Informatika UIN Malang terkhusus Angkatan 2020 “Integer”. Serta teman-teman yang dekat dengan penulis tidak bisa disebutkan satu persatu, terima kasih telah memberikan *support*, motivasi dan bantuannya kepada penulis.
15. Segenap pihak yang telah terlibat secara langsung maupun tidak langsung dalam proses penyusunan skripsi ini.
16. Untuk diri penulis “Nadia Mu’izzatul Ilma” terima kasih sudah bertahan sejauh ini, terima kasih untuk tidak menyerah, terima kasih untuk tidak pernah berhenti mencoba. Tetaplah tegak ya walaupun sering jatuh dan kecewa, jangan pernah puas dengan pencapaian saat ini terus kejar cita-cita, masih banyak hal yang ingin diwujudkan. Terima kasih Nad! *So proud of you*, jangan pernah lelah untuk bertahan dan berkembang ya!

Penulis menyadari dalam penulisan skripsi ini tidak luput dari kesalahan dan masih jauh dari kata sempurna. Maka dari itu, penulis mengharapkan kritik dan saran yang membangun sehingga dapat lebih berkembang.

Malang, 28 Mei 2024

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	ii
<b>HALAMAN PERSETUJUAN</b> .....	iii
<b>HALAMAN PENGESAHAN</b> .....	iv
<b>PERNYATAAN KEASLIAN TULISAN</b> .....	v
<b>HALAMAN MOTTO</b> .....	vi
<b>HALAMAN PERSEMBAHAN</b> .....	vii
<b>KATA PENGANTAR</b> .....	viii
<b>DAFTAR ISI</b> .....	xi
<b>DAFTAR TABEL</b> .....	xiii
<b>DAFTAR GAMBAR</b> .....	xiii
<b>ABSTRAK</b> .....	xx
<b>ABSTRACT</b> .....	xxi
<b>البحث مستخلص</b> .....	xxii
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	5
1.3 Batasan Masalah .....	5
1.4 Tujuan Penelitian .....	6
1.5 Manfaat Penelitian .....	6
<b>BAB II STUDI PUSTAKA</b> .....	7
2.1 Penelitian Terdahulu .....	7
2.2 Serangan Jantung .....	13
2.3 <i>Exploratory Data Analysis</i> .....	16
2.4 Klasifikasi .....	18
2.5 <i>Random Forest</i> .....	20
<b>BAB III METODOLOGI PENELITIAN</b> .....	23
3.1 Desain Sistem .....	23
3.2 Pengumpulan Data .....	24
3.3 <i>Preprocessing</i> .....	27
3.3.1 Pemisahan Data .....	27
3.3.2 <i>Synthetic Minority Oversampling Technique (SMOTE)</i> .....	27
3.1 <i>Exploratory Data Analysis (EDA)</i> .....	28
3.4.1 Visualisasi Distrbusi Variabel Numerikal .....	29
3.2 Penanganan <i>Outliers</i> .....	29
3.3 Implementasi Metode <i>Random Forest</i> .....	32
3.4 Evaluasi .....	34
3.5 Skenario Pengujian .....	36
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	43
4.1 Hasil Uji Coba .....	43
4.1.1 Pengujian <i>Software</i> yang Digunakan .....	43
4.1.2 Pengujian Dengan <i>SMOTE</i> – Penghapusan <i>Outliers</i> .....	44
4.1.2.1 Model A .....	45

4.1.2.2 Model B .....	50
4.1.2.3 Model C .....	54
4.1.3 Pengujian Dengan Penghapusan <i>Outliers</i> – <i>SMOTE</i> .....	61
4.1.3.1 Model A .....	62
4.1.3.2 Model B .....	67
4.1.3.3 Model C .....	71
4.1.4 Pengujian Dengan <i>SMOTE</i> dan Tanpa Penghapusan <i>Outliers</i> .....	78
4.1.4.1 Model A .....	79
4.1.4.2 Model B .....	84
4.1.4.3 Model C .....	88
4.1.5 Pengujian Dengan Penghapusan <i>Outliers</i> dan Tanpa <i>SMOTE</i> .....	95
4.1.5.1 Model A .....	96
4.1.5.2 Model B .....	101
4.1.5.3 Model C .....	105
4.1.6 Pengujian <i>Data Real</i> .....	112
4.1.6.1 Pengujian Dengan <i>SMOTE</i> – Penghapusan <i>Outliers</i> .....	113
4.1.6.2 Pengujian Dengan Penghapusan <i>Outliers</i> - <i>SMOTE</i> .....	118
4.1.6.3 Pengujian Dengan <i>SMOTE</i> dan Tanpa Penghapusan <i>Outliers</i> .	123
4.1.6.4 Pengujian Dengan Penghapusan <i>Outliers</i> dan Tanpa <i>SMOTE</i> ..	128
4.2 Pembahasan.....	135
4.2.1 Pengujian Dengan <i>SMOTE</i> – Penhapusan <i>Outliers</i> .....	135
4.2.2 Pengujian Dengan Penghapusan <i>Outliers</i> – <i>SMOTE</i> .....	137
4.2.1 Pengujian Dengan <i>SMOTE</i> dan Tanpa Penghapusan <i>Outliers</i> .....	140
4.2.2 Pengujian Dengan Penghapusan <i>Outliers</i> dan Tanpa <i>SMOTE</i> .....	142
4.3 Integrasi Penelitian Dalam Tafsir Al-Qur'an.....	144
<b>BAB V KESIMPULAN DAN SARAN</b> .....	147
5.1 Kesimpulan .....	147
5.2 Saran .....	148
<b>DAFTAR PUSTAKA</b>	

## DAFTAR TABEL

Tabel 2. 1 Penelitian terdahulu.....	11
Tabel 3. 1 Detail Fitur Dataset .....	24
Tabel 3. 2 <i>Confusion Matrix</i> .....	35
Tabel 3. 3 Tuning Hyperparameter .....	36
Tabel 3. 4 Skenario Pengujian .....	38
Tabel 4. 1 Hasil Pengujian Model A Kombinasi Fitur 1 .....	45
Tabel 4. 2 <i>Confusion Matrix</i> Pengujian Model A Kombinasi Fitur 1.....	46
Tabel 4. 3 Hasil Evaluasi Pengujian Model A Kombinasi Fitur 1.....	46
Tabel 4. 4 Hasil Pengujian Model A Kombinasi Fitur 2 .....	46
Tabel 4. 5 <i>Confusion Matrix</i> Pengujian Model A Kombinasi Fitur 2.....	47
Tabel 4. 6 Hasil Evaluasi Pengujian Model A Kombinasi Fitur 2.....	47
Tabel 4. 7 Hasil Pengujian Model A Kombinasi Fitur 3 .....	48
Tabel 4. 8 <i>Confusion Matrix</i> Pengujian Model A Kombinasi Fitur 3.....	48
Tabel 4. 9 Hasil Evaluasi Pengujian Model A Kombinasi Fitur 3.....	48
Tabel 4. 10 Hasil Pengujian Model A Semua Fitur .....	49
Tabel 4. 11 <i>Confusion Matrix</i> Pengujian Model A Semua Fitur .....	49
Tabel 4. 12 Hasil Evaluasi Pengujian Model A Semua Fitur .....	50
Tabel 4. 13 Hasil Pengujian Model B Kombinasi Fitur 1.....	50
Tabel 4. 14 <i>Confusion Matrix</i> Pengujian Model B Kombinasi Fitur 1.....	50
Tabel 4. 15 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 1.....	51
Tabel 4. 16 Hasil Pengujian Model B Kombinasi Fitur 2.....	51
Tabel 4. 17 <i>Confusion Matrix</i> Pengujian Model B Kombinasi Fitur 2.....	51
Tabel 4. 18 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 2.....	52
Tabel 4. 19 Hasil Pengujian Model B Kombinasi Fitur 3.....	52
Tabel 4. 20 <i>Confusion Matrix</i> Pengujian Model B Kombinasi Fitur 3.....	53
Tabel 4. 21 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 3.....	53
Tabel 4. 22 Hasil Pengujian Model B Semua Fitur .....	53
Tabel 4. 23 <i>Confusion Matrix</i> Pengujian Model B Semua Fitur .....	54
Tabel 4. 24 Hasil Evaluasi Pengujian Model B Semua Fitur.....	54
Tabel 4. 25 Hasil Pengujian Model C Kombinasi Fitur 1.....	55
Tabel 4. 26 <i>Confusion Matrix</i> Pengujian Model C Kombinasi Fitur 1.....	55
Tabel 4. 27 Hasil Evaluasi Pengujian Model C Kombinasi Fitur 1.....	56
Tabel 4. 28 Hasil Pengujian Model C Kombinasi Fitur 2.....	56
Tabel 4. 29 <i>Confusion Matrix</i> Pengujian Model C Kombinasi Fitur 2.....	56
Tabel 4. 30 Hasil Evaluasi Pengujian Model C Kombinasi Fitur 2.....	57
Tabel 4. 31 Hasil Pengujian Model C Kombinasi Fitur 3.....	57
Tabel 4. 32 <i>Confusion Matrix</i> Pengujian Model C Kombinasi Fitur 3.....	57
Tabel 4. 33 Hasil Evaluasi Pengujian Model C Kombinasi Fitur 3.....	58
Tabel 4. 34 Hasil Pengujian Model C Semua Fitur .....	58
Tabel 4. 35 <i>Confusion Matrix</i> Pengujian Model C Semua Fitur .....	58
Tabel 4. 36 Hasil Evaluasi Pengujian Model C Semua Fitur.....	59
Tabel 4. 37 Hasil Seluruh Pengujian dengan <i>SMOTE</i> - Penghapusan Outliers....	59
Tabel 4. 38 Hasil Pengujian Model A Kombinasi Fitur 1.....	62

Tabel 4. 39	<i>Confusion Matrix</i> Pengujian Model A Kombinasi Fitur 1.....	63
Tabel 4. 40	Hasil Evaluasi Pengujian Model A Kombinasi Fitur 1.....	63
Tabel 4. 41	Hasil Pengujian Model A Kombinasi Fitur 2.....	63
Tabel 4. 42	<i>Confusion Matrix</i> Pengujian Model A Kombinasi Fitur 2.....	64
Tabel 4. 43	Hasil Evaluasi Hasil Pengujian Model A Kombinasi Fitur 2 .....	64
Tabel 4. 44	Hasil Hasil Pengujian Model A Kombinasi Fitur 3 .....	65
Tabel 4. 45	<i>Confusion Matrix</i> Hasil Pengujian Model A Kombinasi Fitur 3 .....	65
Tabel 4. 46	Hasil Evaluasi Hasil Pengujian Model A Kombinasi Fitur 3 .....	65
Tabel 4. 47	Hasil Pengujian Model A Semua Fitur .....	66
Tabel 4. 48	<i>Confusion Matrix</i> Pengujian Model A Semua Fitur .....	66
Tabel 4. 49	Hasil Evaluasi Pengujian Model A Semua Fitur .....	67
Tabel 4. 50	Hasil Pengujian Model B Kombinasi Fitur 1 .....	67
Tabel 4. 51	<i>Confusion Matrix</i> Pengujian Model B Kombinasi Fitur 1 .....	67
Tabel 4. 52	Hasil Evaluasi Pengujian Model B Kombinasi Fitur 1 .....	68
Tabel 4. 53	Hasil Pengujian Model B Kombinasi Fitur 2.....	68
Tabel 4. 54	<i>Confusion Matrix</i> Pengujian Model B Kombinasi Fitur 2.....	68
Tabel 4. 55	Hasil Evaluasi Pengujian Model B Kombinasi Fitur 2.....	69
Tabel 4. 56	Hasil Pengujian Model B Kombinasi Fitur 3.....	69
Tabel 4. 57	<i>Confusion Matrix</i> Pengujian Model B Kombinasi Fitur 3.....	70
Tabel 4. 58	Hasil Evaluasi Pengujian Model B Kombinasi Fitur 3.....	70
Tabel 4. 59	Hasil Pengujian Model B Semua Fitur .....	70
Tabel 4. 60	<i>Confusion Matrix</i> Model B Semua Fitur.....	71
Tabel 4. 61	Hasil Evaluasi Pengujian Model B Semua Fitur.....	71
Tabel 4. 62	Hasil Pengujian Model C Kombinasi Fitur 1.....	72
Tabel 4. 63	<i>Confusion Matrix</i> Pengujian Model C Kombinasi Fitur 1.....	72
Tabel 4. 64	Hasil Evaluasi Pengujian Model C Kombinasi Fitur 1.....	73
Tabel 4. 65	Hasil Pengujian Model C Kombinasi Fitur 2.....	73
Tabel 4. 66	<i>Confusion Mtrix</i> Pengujian Model C Kombinasi Fitur 2 .....	73
Tabel 4. 67	Hasil Evaluasi Model C Kombinasi Fitur 2.....	74
Tabel 4. 68	Hasil Pengujian Model C Kombinasi Fitur 3.....	74
Tabel 4. 69	<i>Confusion Matrix</i> Pengujian Model C Kombinasi Fitur 3.....	74
Tabel 4. 70	Hasil Evaluasi Pengujian Model C Kombinasi Fitur 3.....	75
Tabel 4. 71	Hasil Pengujian Model C Semua Fitur .....	75
Tabel 4. 72	<i>Confusion Matrix</i> Pengujian Model C Semua Fitur .....	75
Tabel 4. 73	Hasil Evaluasi Pengujian Model C Semua Fitur.....	76
Tabel 4. 74	Hasil Keseluruhan Evaluasi Pengujian Dengan Penghapusan Outliers – <i>SMOTE</i> .....	76
Tabel 4. 75	Hasil Pengujian Model A Kombinasi Fitur 1.....	79
Tabel 4. 76	<i>Confusion Matrix</i> Pengujian Model A Kombinasi Fitur 1.....	80
Tabel 4. 77	Hasil Evaluasi Pengujian Model A Kombinasi 1.....	80
Tabel 4. 78	Hasil Evaluasi Pengujian Model A Kombinasi Fitur 2.....	80
Tabel 4. 79	<i>Confusion Matrix</i> Pengujian Model A Kombinasi Fitur 2.....	81
Tabel 4. 80	Hasil Evaluasi Pengujian Model A Kombinasi Fitur 2.....	81
Tabel 4. 81	Hasil Pengujian Model A Kombinasi Fitur 3.....	82
Tabel 4. 82	<i>Confusion Matrix</i> Pengujian Model A Kombinasi Fitur 3.....	82
Tabel 4. 83	Hasil Evaluasi Pengujian Model A Kombinasi Fitur 3.....	82



Tabel 4. 84 Hasil Pengujian Model A Semua Fitur .....	83
Tabel 4. 85 <i>Confusion Matrix</i> Pengujian Model A Semua Fitur .....	83
Tabel 4. 86 Hasil Evaluasi Pengujian Model A Semua Fitur .....	84
Tabel 4. 87 Hasil Pengujian Model B Kombinasi Fitur 1 .....	84
Tabel 4. 88 <i>Confusion Matrix</i> Pengujian Model B Kombinasi Fitur 1 .....	84
Tabel 4. 89 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 1 .....	85
Tabel 4. 90 Hasil Pengujian Model B Kombinasi Fitur 2.....	85
Tabel 4. 91 <i>Confusion Matrix</i> Pengujian Model B Kombinasi Fitur 2.....	85
Tabel 4. 92 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 2.....	86
Tabel 4. 93 Hasil Pengujian Model B Kombinasi Fitur 3.....	86
Tabel 4. 94 <i>Confusion Matrix</i> Pengujian Model B Kombinasi Fitur 3.....	87
Tabel 4. 95 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 3.....	87
Tabel 4. 96 Hasil Pengujian Model B Semua Fitur .....	87
Tabel 4. 97 <i>Confusion Matrix</i> Pengujian Model B Semua Fitur .....	88
Tabel 4. 98 Hasil Evaluasi Model B Semua Fitur.....	88
Tabel 4. 99 Hasil Pengujian Model C Kombinasi Fitur 1 .....	89
Tabel 4. 100 <i>Confusion Matrix</i> Model C Kombinasi Fitur 1 .....	89
Tabel 4. 101 Hasil Evaluasi Model C Kombinasi Fitur 1 .....	90
Tabel 4. 102 Hasil Pengujian Model C Kombinasi Fitur 2.....	90
Tabel 4. 103 <i>Confusion Matrix</i> Pengujian Model C Kombinasi Fitur 2.....	90
Tabel 4. 104 Hasil Evaluasi Model C Kombinasi Fitur 2.....	91
Tabel 4. 105 Hasil Pengujian Model C Kombinasi Fitur 3.....	91
Tabel 4. 106 <i>Confusion Matrix</i> Pengujian Model C Kombinasi Fitur 3.....	91
Tabel 4. 107 Hasil Evaluasi Matrix Pengujian Model C Kombinasi Fitur 3 .....	92
Tabel 4. 108 Hasil Pengujian Model C Semua Fitur .....	92
Tabel 4. 109 <i>Confusion Matrix</i> Pengujian Model C Semua Fitur.....	92
Tabel 4. 110 Hasil Evaluasi Pengujian Model C Semua Fitur.....	93
Tabel 4. 111 Hasil Evaluasi Keseluruhan Pengujian Dengan <i>SMOTE</i> dan Tanpa Penghapusan Outliers .....	93
Tabel 4. 112 Hasil Pengujian Model A Kombinasi Fitur 1 .....	96
Tabel 4. 113 <i>Confusion Matrix</i> Pengujian Model A Kombinasi Fitur 1.....	97
Tabel 4. 114 Hasil Evaluasi Pengujian Model A Kombinasi Fitur 1 .....	97
Tabel 4. 115 Hasil Pengujian Model A Kombinasi Fitur 2 .....	97
Tabel 4. 116 <i>Confusion Matrix</i> Model A Kombinasi Fitur 2.....	98
Tabel 4. 117 Hasil Evaluasi Model A Kombinasi Fitur 2.....	98
Tabel 4. 118 Hasil Pengujian Model A Kombinasi Fitur 3.....	99
Tabel 4. 119 <i>Confusion Matrix</i> Pengujian Model A Kombinasi Fitur 3.....	99
Tabel 4. 120 Hasil Evaluasi Pengujian Model A Kombinasi Fitur 3.....	99
Tabel 4. 121 Hasil Pengujian Model A Semua Fitur .....	100
Tabel 4. 122 <i>Confusion Matrix</i> Pengujian Model A Semua Fitur .....	100
Tabel 4. 123 Hasil Evaluasi Model A Semua Fitur .....	101
Tabel 4. 124 Hasil Pengujian Model B Kombinasi Fitur 1 .....	101
Tabel 4. 125 <i>Confusion Matrix</i> Model B Kombinasi Fitur 1.....	101
Tabel 4. 126 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 1 .....	102
Tabel 4. 127 Hasil Pengujian Model B Kombinasi Fitur 2.....	102
Tabel 4. 128 <i>Confusion Matrix</i> Model B Kombinasi Fitur 2.....	102

Tabel 4. 129 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 2.....	103
Tabel 4. 130 Hasil Pengujian Model B Kombinasi Fitur 3.....	103
Tabel 4. 131 <i>Confusion Matrix</i> Pengujian Model B Kombinasi Fitur 3.....	104
Tabel 4. 132 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 3.....	104
Tabel 4. 133 Hasil Pengujian Model B Semua Fitur .....	104
Tabel 4. 134 <i>Confusion Matrix</i> Model B Semua Fitur.....	105
Tabel 4. 135 Hasil Evaluasi Pengujian Model B Semua Fitur.....	105
Tabel 4. 136 Hasil Pengujian Model C Kombinasi Fitur 1 .....	106
Tabel 4. 137 <i>Confusion Matrix</i> Pengujian Model C Kombinasi Fitur 1 .....	106
Tabel 4. 138 Hasil Evaluasi Pengujian Model C Kombinasi Fitur 1 .....	107
Tabel 4. 139 Hasil Pengujian Model C Kombinasi Fitur 2.....	107
Tabel 4. 140 <i>Confusion Matrix</i> Pengujian Model C Kombinasi Fitur 2.....	107
Tabel 4. 141 Hasil Evaluasi Pengujian Model C Kombinasi Fitur 2.....	108
Tabel 4. 142 Hasil Pengujian Model C Kombinasi Fitur 3.....	108
Tabel 4. 143 <i>Confusion Matrix</i> Pengujian Model C Kombinasi Fitur 3.....	108
Tabel 4. 144 Hasil Evaluasi Pengujian Model C Kombinasi Fitur 3.....	109
Tabel 4. 145 Hasil Pengujian Model C Semua Fitur .....	109
Tabel 4. 146 <i>Confusion Matrix</i> Pengujian Model C Semua Fitur.....	109
Tabel 4. 147 Hasil Evaluasi Pengujian Model C Semua Fitur.....	110
Tabel 4. 148 Hasil Evaluasi Seluruh Pengujian Dengan Penghapusan Outliers dan Tanpa <i>SMOTE</i> .....	110
Tabel 4. 149 Hasil Pengujian <i>SMOTE</i> - Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 1 .....	113
Tabel 4. 150 <i>Confusion Matrix</i> Pengujian <i>SMOTE</i> - Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 1 .....	114
Tabel 4. 151 Hasil Evaluasi Pengujian <i>SMOTE</i> - Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 1 .....	114
Tabel 4. 152 Hasil Pengujian <i>SMOTE</i> - Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 2 .....	115
Tabel 4. 153 <i>Confusion Matrix</i> Pengujian <i>SMOTE</i> - Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 2 .....	115
Tabel 4. 154 Hasil Evaluasi Pengujian <i>SMOTE</i> - Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 2 .....	116
Tabel 4. 155 Hasil Pengujian <i>SMOTE</i> - Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 3 .....	116
Tabel 4. 156 <i>Confusion Matrix</i> Pengujian <i>SMOTE</i> - Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 3 .....	116
Tabel 4. 157 Hasil Evaluasi Pengujian <i>SMOTE</i> - Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 3.....	117
Tabel 4. 158 Hasil Pengujian <i>SMOTE</i> - Penghapusan Outliers Data <i>Real</i> Semua Fitur.....	117
Tabel 4. 159 <i>Confusion Matrix</i> Pengujian <i>SMOTE</i> - Penghapusan Outliers Data <i>Real</i> Semua Fitur .....	117
Tabel 4. 160 Hasil Evaluasi Pengujian <i>SMOTE</i> - Penghapusan Outliers Data <i>Real</i> Semua Fitur .....	118

Tabel 4. 161 Hasil Pengujian Penghapusan Outliers - <i>SMOTE Data Real</i> Kombinasi Fitur 1 .....	119
Tabel 4. 162 <i>Confusion Matrix</i> Pengujian Penghapusan Outliers - <i>SMOTE</i> <i>Data Real</i> Kombinasi Fitur 1 .....	119
Tabel 4. 163 Hasil Evaluasi Pengujian Penghapusan Outliers - <i>SMOTE Data</i> <i>Real</i> Kombinasi Fitur 1 .....	120
Tabel 4. 164 Hasil Pengujian Penghapusan Outliers - <i>SMOTE Data Real</i> Kombinasi Fitur 2 .....	120
Tabel 4. 165 <i>Confusion Matrix</i> Pengujian Penghapusan Outliers - <i>SMOTE</i> <i>Data Real</i> Kombinasi Fitur 2 .....	120
Tabel 4. 166 Hasil Evaluasi Pengujian Penghapusan Outliers - <i>SMOTE Data</i> <i>Real</i> Kombinasi Fitur 2 .....	121
Tabel 4. 167 Hasil Pengujian Penghapusan Outliers - <i>SMOTE Data Real</i> Kombinasi Fitur 3 .....	121
Tabel 4. 168 <i>Confusion Matrix</i> Pengujian Penghapusan Outliers - <i>SMOTE</i> <i>Data Real</i> Kombinasi Fitur 3 .....	121
Tabel 4. 169 Hasil Evaluasi Pengujian Penghapusan Outliers - <i>SMOTE Data</i> <i>Real</i> Kombinasi Fitur 3 .....	122
Tabel 4. 170 Pengujian Penghapusan Outliers - <i>SMOTE Data Real</i> Semua Fitur .....	122
Tabel 4. 171 <i>Confusion Matrix</i> Pengujian Penghapusan Outliers - <i>SMOTE</i> <i>Data Real</i> Semua Fitur .....	122
Tabel 4. 172 Hasil Evaluasi Pengujian Penghapusan Outliers - <i>SMOTE Data</i> <i>Real</i> Semua Fitur .....	123
Tabel 4. 173 Hasil Pengujian Dengan <i>SMOTE</i> Kombinasi Fitur 1 .....	124
Tabel 4. 174 <i>Confusion Matrix</i> Pengujian Dengan <i>SMOTE Data Real</i> Kombinasi Fitur 1 .....	124
Tabel 4. 175 Hasil Evaluasi Pengujian Dengan <i>SMOTE Data Real</i> Kombinasi Fitur 1 .....	125
Tabel 4. 176 Hasil Pengujian Dengan <i>SMOTE Data Real</i> Kombinasi Fitur 2 ....	125
Tabel 4. 177 Hasil Evaluasi Pengujian Dengan <i>Data Real SMOTE</i> Kombinasi Fitur 2 .....	125
Tabel 4. 178 Hasil Evaluasi Pengujian Dengan <i>SMOTE Data Real</i> Kombinasi Fitur 2 .....	126
Tabel 4. 179 Hasil Pengujian Dengan <i>SMOTE Data Real</i> Kombinasi Fitur 3 ....	126
Tabel 4. 180 <i>Confusion Matrix</i> Pengujian Dengan <i>SMOTE Data Real</i> Kombinasi Fitur 3 .....	127
Tabel 4. 181 Hasil Evaluasi Pengujian Dengan <i>SMOTE Data Real</i> Kombinasi Fitur 3 .....	127
Tabel 4. 182 Hasil Pengujian Dengan <i>SMOTE Data Real</i> Semua Fitur .....	127
Tabel 4. 183 <i>Confusion Matrix</i> Pengujian Dengan <i>SMOTE Data Real</i> Semua Fitur .....	128
Tabel 4. 184 Hasil Evaluasi Pengujian Dengan <i>SMOTE Data Real</i> Semua Fitur .....	128
Tabel 4. 185 Hasil Pengujian Dengan Penghapusan Outliers <i>Data Real</i> Kombinasi Fitur 1 .....	129

Tabel 4. 186	<i>Confusion Matrix</i> Pengujian Dengan Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 1 .....	129
Tabel 4. 187	Hasil Evaluasi Pengujian Dengan Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 1 .....	130
Tabel 4. 188	Hasil Pengujian Dengan Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 2 .....	130
Tabel 4. 189	<i>Confusion Matrix</i> Pengujian Dengan Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 2 .....	130
Tabel 4. 190	Hasil Evaluasi Pengujian Dengan Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 2 .....	131
Tabel 4. 191	Hasil Pengujian Dengan Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 3 .....	131
Tabel 4. 192	<i>Confusion Matrix</i> Pengujian Dengan Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 3 .....	132
Tabel 4. 193	Hasil Evaluasi Pengujian Dengan Penghapusan Outliers Data <i>Real</i> Kombinasi Fitur 3 .....	132
Tabel 4. 194	Hasil Pengujian Dengan Penghapusan Outliers Data <i>Real</i> Semua Fitur .....	132
Tabel 4. 195	<i>Confusion Matrix</i> Pengujian Dengan Penghapusan Outliers Data <i>Real</i> Semua Fitur .....	133
Tabel 4. 196	Hasil Evaluasi Pengujian Dengan Penghapusan Outliers Data <i>Real</i> Semua Fitur .....	133
Tabel 4. 197	Hasil Evaluasi Keseluruhan Pengujian Data <i>Real</i> .....	134
Tabel 4. 198	Hasil Terbaik Dari Pengujian Dengan <i>SMOTE</i> – Penghapusan Outliers.....	137
Tabel 4. 199	Hasil Terbaik Dari Pengujian Dengan Penghapusan Outliers – <i>SMOTE</i> .....	140
Tabel 4. 200	Hasil Terbaik Dari Pengujian Dengan <i>SMOTE</i> – Penghapusan Outliers.....	142
Tabel 4. 201	Hasil Terbaik Dari Pengujian Dengan Penghapusan Outliers – <i>SMOTE</i> .....	143
Tabel 4. 202	Hasil Terbaik Dari Setiap Pengujian.....	144

## DAFTAR GAMBAR

Gambar 2. 1 Prevalensi Penyakit Jantung Tahun 2018.....	15
Gambar 2. 2 Grafik Penyakit Jantung di Indonesia .....	15
Gambar 2. 3 Alur Metode Random Forest.....	20
Gambar 3. 1 Desain Sistem.....	24
Gambar 4. 1 Hasil Pengujian Software .....	44
Gambar 4. 2 Fitur Penting Pengujian Dengan <i>SMOTE</i> – Penghapusan Outliers..	45
Gambar 4. 3 Grafik Evaluasi Matrix Pengujian Dengan <i>SMOTE</i> – Penghapusan Outliers .....	60
Gambar 4. 4 Fitur Penting Pengujian Dengan Penghapusan Outliers – <i>SMOTE</i> ..	62
Gambar 4. 5 Grafik Evaluasi Matrix Pengujian Dengan Penghapusan Outliers – <i>SMOTE</i> .....	77
Gambar 4. 6 Fitur Penting Pengujian Dengan <i>SMOTE</i> dan Tanpa Penghapusan Outliers .....	79
Gambar 4. 7 Grafik Evaluasi Matrix Pengujian Dengan <i>SMOTE</i> dan Tanpa Penghapusan Outliers .....	94
Gambar 4. 8 Fitur Penting Pengujian Dengan Penghapusan Outliers dan Tanpa <i>SMOTE</i> .....	96
Gambar 4. 9 Grafik Evaluasi Pengujian Dengan Penghapusan Outliers dan Tanpa <i>SMOTE</i> .....	111
Gambar 4. 10 Fitur Penting Data <i>Real</i> Pengujian Dengan <i>SMOTE</i> – Penghapusan Outliers .....	113
Gambar 4. 11 Fitur Penting Data <i>Real</i> Pengujian Dengan Penghapusan Outliers - <i>SMOTE</i> .....	118
Gambar 4. 12 Fitur Penting Data <i>Real</i> Pengujian Dengan <i>SMOTE</i> dan Tanpa Penghapusan Outliers .....	124
Gambar 4. 13 Fitur Penting Data <i>Real</i> Pengujian Dengan Penghapusan Outliers dan Tanpa <i>SMOTE</i> .....	129
Gambar 4. 14 Grafik Perbandingan Sebelum dan Setelah <i>SMOTE</i> .....	135
Gambar 4. 15 Outliers yang Terdeteksi .....	136
Gambar 4. 16 Setelah Proses Penghapusan Outliers.....	136
Gambar 4. 17 Outliers yang Terdeteksi .....	138
Gambar 4. 18 Setelah Proses Penanganan Outliers .....	138
Gambar 4. 19 Grafik Data Sebelum <i>SMOTE</i> .....	139
Gambar 4. 20 Grafik Data Setelah <i>SMOTE</i> .....	139
Gambar 4. 21 Grafik Data Sebelum dan Setelah <i>SMOTE</i> .....	141
Gambar 4. 22 Outliers yang Terdeteksi .....	142
Gambar 4. 23 Setelah Penghapusan Outliers .....	143

## ABSTRAK

Ilma, Nadia Mu'izzatul. 2024. **Implementasi Algoritma *Random Forest* Untuk Klasifikasi Risiko Serangan Jantung. Skripsi.** Program Studi Teknik Informatika, Fakultas Sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Prof. Dr. Suhartono, M.Kom. (II) Dr. Totok Chamidy, M.Kom.

**Kata Kunci:** Klasifikasi, Risiko Serangan Jantung, *Random Forest*, *SMOTE*

Serangan jantung, yang merupakan penyebab kematian utama di Indonesia dan di dunia, dipicu oleh berbagai faktor seperti penyakit jantung koroner, tekanan darah tinggi, dan gaya hidup tidak sehat. Berdasarkan *Global Burden of Disease* dan *Institute for Health and Metrics and Evaluation (IHME)* pada tahun 2014-2019 penyakit jantung menjadi penyebab kematian tertinggi di Indonesia. Berdasarkan Laporan Status Global Penyakit Kardiovaskular pada tahun 2021 oleh *World Health Organization (WHO)* penyakit serangan jantung merupakan penyebab kematian kardiovaskular nomor satu di dunia, dengan 16,17% dari total kematian kardiovaskular. Penelitian ini bertujuan mengkaji implementasi algoritma *Random Forest* untuk mengklasifikasikan risiko serangan jantung. Dalam penelitian ini, data dianalisis menggunakan berbagai skenario pengujian, yaitu kombinasi penghapusan *outliers* dan teknik *oversampling SMOTE*. Pada masing-masing pengujian dilakukan dengan menggunakan 3 model pembagian data, yaitu model A dengan perbandingan data 90:10, model B dengan perbandingan data 80:20, dan model C dengan perbandingan data 75:25. Untuk masing-masing model dilakukan pengujian dengan 4 kelompok kombinasi data berdasarkan fitur penting. Hasil pengujian menunjukkan bahwa kombinasi penghapusan *outliers* dan *SMOTE* pada model A memberikan akurasi tertinggi sebesar 93%. Evaluasi lebih lanjut mengindikasikan bahwa algoritma *Random Forest* memiliki performa yang sangat baik dengan nilai akurasi, *precision*, *recall*, dan *f1-score* sebesar 100% dalam pengujian pada data *Real*. Temuan ini menunjukkan bahwa algoritma *Random Forest* efektif dalam mengklasifikasikan populasi yang berisiko serangan jantung, memberikan potensi untuk turut dalam upaya pencegahan dan penanganan dini penyakit jantung.



## ABSTRACT

Ilma, Nadia Mu'izzatul. 2024. *Implementation of Random Forest Algorithm for Heart Attack Risk Classification*. Department of Informatics Engineering Faculty of Science and Technology Islamic State University of Maulana Malik Ibrahim Malang. Supervisors: (I) Prof. Dr. Suhartono, M.Kom. (II) Dr. Totok Chamidy, M.Kom.

Heart attacks, which are the leading cause of death in Indonesia and worldwide, are triggered by various factors such as coronary heart disease, high blood pressure, and unhealthy lifestyles. Based on the Global Burden of Disease and the Institute for Health and Metrics and Evaluation (IHME) in 2014-2019, heart disease is the highest cause of death in Indonesia. Based on the Global Status Report on Cardiovascular Disease in 2021 by the World Health Organization (WHO), heart attack disease is the number one cause of cardiovascular death in the world, with 16.17% of total cardiovascular deaths. This study aims to examine the implementation of the Random Forest algorithm to classify the risk of heart attack. In this study, the data were analyzed using various testing scenarios, namely a combination of outlier removal and *SMOTE* oversampling techniques. Each test was conducted using 3 data division models, namely model A with a data ratio of 90:10, model B with a data ratio of 80:20, and model C with a data ratio of 75:25. For each model, 4 groups of data combinations based on important features were tested. The test results showed that the combination of outlier removal and *SMOTE* in model A gave the highest accuracy of 93%. Further evaluation indicated that the Random Forest algorithm performed very well with accuracy, precision, recall, and f1-score values of 100% in testing on *Real* data. These findings suggest that the Random Forest algorithm is effective in classifying populations at risk of heart attack, providing the potential to contribute to the prevention and early treatment of heart disease.

**Keywords:** *Classification, Heart Attack Risk, Random Forest, SMOTE*

## البحث مستخلص

العلم، نادي معزة. ٢٠٢٤. تنفيذ خوارزمية **Random Forest** لتصنيف مخاطر النوبات القلبية. البحث الجامعي. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، الجامعة الإسلامية الحكومية مولانا مالك إبراهيم مالانج. المشرف: (١) بروفييسور الدكتور سوهارتونو الماجستير (٢) الدكتور توتوك حاميدي الماجستير

**الكلمات الرئيسية:** التصنيف، خطر الإصابة بالنوبات القلبية، *SMOTE*، *Random Forest*

تتنجم النوبات القلبية، وهي السبب الرئيسي للوفاة في إندونيسيا والعالم، عن عوامل مختلفة مثل أمراض القلب التاجية وارتفاع ضغط الدم وأنماط الحياة غير الصحية. واستنادًا إلى العبء العالمي للأمراض ومعهد الصحة والقياسات والتقييم (IHME) في الفترة 2014-2019، فإن أمراض القلب هي السبب الرئيسي للوفاة في إندونيسيا. استنادًا إلى تقرير الحالة العالمية لأمراض القلب والأوعية الدموية في عام 2021 الصادر عن منظمة الصحة العالمية، فإن مرض النوبات القلبية هو السبب الأول لوفيات القلب والأوعية الدموية في العالم، حيث تبلغ نسبته 16.17% من إجمالي وفيات القلب والأوعية الدموية. يهدف هذا البحث إلى دراسة تطبيق خوارزمية الغابة العشوائية لتصنيف خطر الإصابة بالنوبات القلبية. في هذه الدراسة، تم تحليل البيانات باستخدام سيناريوهات اختبار مختلفة، وهي مزيج من تقنيات إزالة القيم المتطرفة وتقنيات أخذ العينات الزائدة من *SMOTE*. تم إجراء كل اختبار باستخدام 3 نماذج لتقسيم البيانات، وهي النموذج (أ) بنسبة بيانات 90:10، والنموذج (ب) بنسبة بيانات 80:20، والنموذج (ج) بنسبة بيانات 75:25. لكل نموذج، تم اختبار 4 مجموعات من مجموعات البيانات بناءً على السمات المهمة. أظهرت نتائج الاختبار أن الجمع بين إزالة القيم المتطرفة و *SMOTE* في النموذج (أ) بنسبة بيانات 90:10 أعطى أعلى دقة بنسبة 93%. أشار المزيد من التقييم إلى أن خوارزمية الغابة العشوائية كان أداءها جيدًا جدًا حيث بلغت قيم الدقة والدقة والاستدعاء ودرجة  $f1$  100% في الاختبار على بيانات حقيقية. تشير هذه النتائج إلى أن خوارزمية الغابة العشوائية فعالة في تصنيف الفئات السكانية المعرضة لخطر الإصابة بالنوبات القلبية، مما يوفر إمكانية المساهمة في الوقاية والعلاج المبكر لأمراض القلب.

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Serangan jantung, atau yang dikenal secara medis sebagai infark miokard, merupakan suatu kondisi dimana aliran darah ke jantung sangat berkurang atau tersumbat dikarenakan pembuluh arteri mengalami penyempitan karena adanya sumbatan sehingga mengakibatkan suplai oksigen dalam darah tidak sesuai (Amrullah et al., 2022). Sebagaimana firman Allah pada surah Al-Haqqah ayat 46.

ثُمَّ لَقَطَعْنَا مِنْهُ الْوَتِينَ

“Kemudian benar-benar kami potong urat tali jantungnya” (Q.S. Al-Haqqah: 46).

Pembuluh nadi, dalam hal ini merupakan pembuluh darah besar atau dalam dunia medis disebut aorta yang tekanannya langsung dari kontraksi jantung dan volume darah yang besar, tersumbat atau terpotongnya aorta dapat mengakibatkan syok bahkan kematian (Tafsir Ibnu Katsir, n.d.). Ayat diatas menunjukkan pentingnya pemahaman tentang fungsi dan kerentanan jantung dalam konteks risiko serangan jantung. Dalam penelitian ini, ayat di atas dihubungkan dengan pemahaman akan akibat yang cukup serius dari gangguan pada jantung.

أَلَا وَإِنَّ فِي الْجَسَدِ مُضغَةً إِذَا صَلَحَتْ صَلَحَ الْجَسَدُ كُلُّهُ ، وَإِذَا فَسَدَتْ فَسَدَ الْجَسَدُ كُلُّهُ . أَلَا وَهِيَ  
الْقَلْبُ

*“Ingatlah bahwa di dalam jasad itu ada segumpal daging. Jika ia baik, maka baik pula seluruh jasad. Jika ia rusak, maka rusak pula seluruh jasad. Ketauhilah bahwa ia adalah qalbun (jantung)” (HR Bukhari No 52 dan Muslim No 1599).*

Menurut Ibnu Katsir hubungan antara surah Al-Haqqah ayat 46, dan hadits riwayat Bukhari dan Muslim diatas dengan penelitian ini terletak pada penekanan terhadap pentingnya Kesehatan jantung. Surah Al-Haqqah ayat 46 menggambarkan konsekuensi fatal dari kerusakan jantung. Dalam bahasa arab jantung disebut sebagai **الْقَلْبُ**, seperti pada hadits diatas yang menekankan peran sentral jantung dalam kesehatan tubuh secara keseluruhan.

Penyakit jantung terutama serangan jantung masih menjadi salah satu penyebab kematian terbanyak di Indonesia. Berdasarkan *Global Burden of Disease* dan *Institute for Health and Metrics and Evaluation (IHME)* 2014-2019 penyakit jantung menjadi penyebab kematian tertinggi di Indonesia. Berdasarkan Laporan Status Global Penyakit Kardiovaskular pada tahun 2021 oleh *World Health Organization (WHO)* penyakit serangan jantung merupakan penyebab kematian kardiovaskular nomor satu di dunia, dengan 16,17% dari total kematian kardiovaskular dengan lebih dari 17 juta kematian pertahun. Berdasarkan data Riset Kesehatan Dasar (Riskesdas) tahun 2018, angka kejadian penyakit jantung semakin meningkat dari tahun ke tahun. Menurut Direktur Pencegahan dan Pengendalian Penyakit Tidak Menular Kementerian Kesehatan setidaknya, 15 dari 1000 orang di Indonesia menderita penyakit jantung. Dengan jumlah kematian yang disebabkan serangan jantung sebesar 14,38% (Kementerian Kesehatan, 2022). Menurut Statistik Kesehatan Indonesia tahun 2021 oleh Badan Pusat

Statistik (BPS) penyakit serangan jantung berada di urutan ke-2 penyebab kematian terbanyak di Indonesia dengan 15,1% dari total kematian.

Penyakit jantung ini, termasuk Serangan jantung disebabkan oleh beberapa faktor yang dapat memicunya. Faktor-faktor seperti adanya penyakit jantung koroner, jenis kelamin, tekanan darah, pola hidup yang tidak sehat, konsumsi rokok dan alkohol, kolesterol tinggi, diabetes terbukti dapat memicu terjadinya serangan jantung (Rohayati & Widani, 2020). Tindakan pencegahan yang efektif untuk mengatasi faktor risiko ini sangat penting dalam mengurangi angka kejadian penyakit jantung terutama serangan jantung.

Kementerian Kesehatan Republik Indonesia melakukan beberapa upaya untuk mencegah masalah penyakit jantung di Indonesia, seperti melakukan penguatan pada layanan primer melalui edukasi penduduk, pencegahan primer dengan imunisasi rutin menjadi 14 antigen dan perluasan cakupan di seluruh Indonesia, pencegahan sekunder dengan dilakukannya skrining 14 penyakit penyebab kematian tertinggi termasuk penyakit jantung pada tiap sasaran usia, serta meningkatkan kapasitas serta kapabilitas layanan primer (Kementerian Kesehatan, 2022).

Selain itu, pencegahan risiko penyakit serangan jantung ini harus didukung dengan kesadaran pribadi setiap individu akan pentingnya menjaga pola hidup yang sehat, serta menjauhi hal-hal yang dapat memicu timbulnya faktor-faktor yang menyebabkan terjadinya serangan jantung (Gadjah Mada, 2021). Untuk melaksanakan strategi ini dengan efektif, tentunya diperlukan pemahaman yang

lebih mendalam terkait faktor-faktor risiko dan identifikasi terhadap individu yang memiliki risiko tinggi.

Pemanfaatan kemajuan teknologi menjadi sangat relevan di masa sekarang ini untuk turut mencegah terjadinya risiko serangan jantung. Kemajuan teknologi terutama dalam analisis data salah satunya data *mining* dapat dimanfaatkan sebagai upaya pencegahan terjadinya risiko ini. Klasifikasi risiko serangan jantung berdasarkan faktor-faktor tertentu menggunakan algoritma *machine learning* dapat membantu dalam upaya pencegahan risiko serangan jantung (Al Azhima et al., 2022). Diantara pendekatan yang dapat digunakan yaitu dengan menerapkan algoritma *Random Forest*.

Algoritma *Random Forest* adalah suatu metode *ensemble learning* dalam *machine learning* yang memanfaatkan banyak pohon keputusan (*decision tree*) untuk membuat klasifikasi yang lebih akurat dan kuat. *Random Forest* merupakan salah satu algoritma yang sangat populer dan kuat dalam *machine learning* karena dinilai mampu untuk menghasilkan klasifikasi yang akurat, mengatasi overfitting, dan memberikan wawasan tentang pentingnya fitur (Jin et al., 2020). Seperti pada penelitian (Sumwiza et al., 2023) yang berjudul “*Enhanced Cardiovascular Disease Prediction Model Using Random Forest Algorithm*”, penelitian ini menghasilkan akurasi yang baik untuk metode *Random Forest* yaitu sebesar 99%. Pada penelitian (Kadhim & Radhi, 2023) yang berjudul “*Heart Disease Classification Using Optimized Machine Learning Algorithms*”, penelitian ini menghasilkan akurasi terbaik yang diperoleh dengan menggunakan algoritma *Random Forest* sebesar 95,4%. Serta pada penelitian (Radhika & Thomas George,



2021) algoritma *Random Forest* menghasilkan akurasi dengan menggunakan *Scikit-Learn* sebesar 84,5%, 90,7%, dan menggunakan *Weka* sebesar 81%.

Dengan menggabungkan *dataset*, informasi tentang faktor-faktor pemicu risiko serangan jantung, dan algoritma *Random Forest* dapat membantu dalam membuat klasifikasi yang lebih akurat tentang risiko serangan jantung. Dengan demikian, penelitian ini bertujuan untuk mengaplikasikan algoritma *Random Forest* pada dataset yang relevan untuk mengklasifikasikan risiko serangan jantung, dengan harapan hasil dari penelitian ini dapat memberikan kontribusi pada upaya pencegahan risiko serangan jantung.

## **1.2 Rumusan Masalah**

Bagaimana mengklasifikasi populasi yang berisiko serangan jantung?

## **1.3 Batasan Masalah**

1. Menerapkan metode *Random Forest Classifier*
2. Memakai data publik yaitu *Heart Attack Analysis & Prediction Dataset*
3. Menggunakan seluruh atribut pada data *Heart Attack Analysis & Prediction Dataset*.
4. Data dibagi menjadi 3 Model untuk pengujian
5. Menggunakan bahasa pemrograman *Python*
6. Menggunakan *software Google Colab*

#### **1.4 Tujuan Penelitian**

Tujuan utama penelitian ini untuk mengklasifikasikan risiko serangan jantung dengan kelas target 0 (tidak berisiko serangan jantung) dan 1 (berisiko serangan jantung) berdasarkan data *Heart Attack Analysis & Prediction Dataset*.

#### **1.5 Manfaat Penelitian**

1. Penelitian ini dapat memberikan manfaat untuk klasifikasi risiko serangan jantung. Hal ini dapat membantu dalam upaya mencegah risiko serangan jantung.
2. Memberikan wawasan tentang metode *Random Forest Classifier*
3. Hasil dari penelitian ini dapat digunakan sebagai landasan pada penelitian selanjutnya.

## BAB II STUDI PUSTAKA

### 2.1 Penelitian Terdahulu

Dalam penelitian (Qomariyah et al., 2018) tentang Penerapan Metode *Artificial Neural Network* untuk Mendeteksi Serangan Jantung di RS Awal Bros Bekasi. Penelitian ini membahas penerapan metode *Artificial Neural Network* (ANN) dengan algoritma *Back Propagation Neural Network* (BPNN) untuk mendeteksi serangan jantung pada pasien di RS Awal Bros Bekasi. Data yang digunakan dalam penelitian ini adalah data rekam medis pasien dengan kode *diagnose* “I” (*suspect* penyakit jantung) dan pasien yang berobat ke poli jantung di RS Awal Bros Bekasi. Hasil penelitian menunjukkan akurasi klasifikasi sebesar 96,47%. Perbedaan penelitian tersebut dengan penelitian ini adalah metode yang digunakan berbeda, yaitu metode *Artificial Neural Network*. Selain itu data yang digunakan pada penelitian Qomariyah (2018) adalah data *private* yaitu data rekam medis dengan kode *diagnose* “I” (*suspect* penyakit jantung) dan pasien yang berobat ke poli jantung RS Awal Bros Bekasi.

Penelitian (Sumwiza et al., 2023) yang berjudul “*Enhanced Cardiovascular Disease Prediction Model Using Random Forest Algorithm*”. Penelitian ini mengusulkan metode *ensemble* berdasarkan algoritma *Random Forest* untuk meningkatkan akurasi klasifikasi penyakit kardiovaskular. Data yang digunakan merupakan data publik yaitu *Heart Disease Dataset* dengan jumlah 1025 dan 14 atribut. Pada tahap *preprocessing* data dilakukan penghapusan *outlier* untuk meningkatkan representasi data fitur terkait penyakit. Hasil yang

didapatkan pada penelitian ini mencapai akurasi 99%. Perbedaan penelitian tersebut dengan penelitian ini terdapat pada data yang digunakan. Penelitian oleh Sumwiza (2023) menggunakan data *Heart Disease Dataset*, sedangkan peneliti menggunakan data *Heart Attack Analysis & Prediction Dataset*. Selain itu terdapat perbedaan pada teknik *SMOTE* yang peneliti gunakan untuk penelitian ini.

Penelitian (Scarano et al., 2023) yang berjudul “*Injury Severity Prediction of Cyclist Crashes Using Random Forests and Random Parameters Logit Models*”. Penelitian ini menggunakan dataset yang terdiri dari 72.363 kecelakaan pengendara sepeda yang terjadi di Inggris dari tahun 2016 hingga tahun 2018 yang diperoleh dari dataset STATS 19 yang disediakan oleh Departemen Transportasi di Inggris. Penelitian ini menggunakan *Random Forest* dan *Random Parameters Logit Model*. Algoritma *Random Forest* menunjukkan kinerja prediktif yang unggul dalam *F-measure* dan *G-mean* yang menunjukkan keakuratannya dalam memprediksi kecelakaan pengendara sepeda yang fatal dan parah. *Random Parameters Logit Model* menunjukkan heterogenies signifikan yang tidak teramati pada data dan memiliki kecocokan yang baik. Perbedaan penelitian tersebut dengan penelitian yang dilakukan oleh peneliti terdapat pada metode yang digunakan, yaitu *Random Forest* dan *Random Parameters Logic Model*, objek penelitian yang digunakan adalah kecelakaan sepeda di Inggris, serta pada data yang digunakan, yaitu *dataset* STATS19 dari tau 2016-2018 dari Departemen Transportasi Inggris.

Penelitian (Depari et al., 2022) tentang Perbandingan Model *Decision Tree*, *Naïve Bayes*, dan *Random Forest* untuk Klasifikasi Penyakit Jantung. Dalam penelitiannya ini membandingkan tiga metode yaitu *Decision Tree*, *Naïve Bayes*, dan *Random Forest* dengan tujuan untuk menganalisis data, membandingkan metode klasifikasi, dan menentukan metode terbaik untuk klasifikasi penyakit jantung. Data yang digunakan adalah *Personal Key Indicators of Heart Disease Dataset*. Hasil penelitian menunjukkan akurasi metode *Decision Tree* adalah 71%, *Naïve Bayes* adalah 72%, dan *Random Forest* adalah 75%. Perbedaan penelitian tersebut dengan penelitian yang dilakukan oleh peneliti terdapat pada data yang digunakan yaitu *Personal Key Indicators of Heart Disease Dataset*, serta adanya perbandingan beberapa algoritma *machine learning* yang digunakan, yaitu *Decision Tree*, *Naïve Bayes*, dan *Random Forest*.

Penelitian (Nugraha, 2021) tentang Klasifikasi Penyakit Jantung *Cardiovascular* Menggunakan Model Algoritma Klasifikasi. Penelitian ini dilakukan untuk membandingkan beberapa algoritma klasifikasi yang paling akurat untuk mengklasifikasi penyakit jantung *cardiovascular*. Penelitian ini menggunakan beberapa algoritma klasifikasi pada *machine learning* seperti *Random Forest*, *Support Vector Machine (SVM)*, *Gradient Boosting Machine*, *XGBOOST*, dan *Light GBM* dengan menggunakan bahasa pemrograman *Python3*. Hasil penelitian ini menunjukkan akurasi *Random Forest* adalah 77%, *SVM* adalah 80%, *Gradient Boosting Machine* adalah 78%, *XGBOOST* adalah 80%, dan *Light GBM* adalah 68%. Perbedaan penelitian tersebut dengan penelitian yang dilakukan oleh peneliti terdapat pada data yang digunakan yaitu

*Heart Failure Prediction Dataset*, serta adanya perbandingan beberapa algoritma *machine learning* yang digunakan, yaitu *Support Vector Machine* (SVM), *Gradient Boosting Machine*, *XGBOOST*, dan *Light GBM*.

Penelitian (Annisa, 2019) tentang Analisis Komparasi Algoritma Klasifikasi Data *Mining* untuk Prediksi Penderita Penyakit Jantung. Penelitian ini membandingkan beberapa algoritma klasifikasi seperti *Decision Tree*, *Naïve Bayes*, *k-Nearest Neighbour*, *Random Forest*, dan *Decision Stump* untuk klasifikasi penderita jantung dengan menggunakan data publik penderita penyakit jantung laki-laki dari *UCI Repository*. Hasil penelitian ini menunjukkan akurasi *Decision Tree* adalah 76%, *Naïve Bayes* adalah 61%, *k-Nearest Neighbour* adalah 79%, *Random Forest* adalah 80%, dan *Decision Stump* adalah 79%. Perbedaan penelitian tersebut dengan penelitian yang dilakukan oleh peneliti terdapat pada data yang digunakan yaitu data penderita penyakit jantung laki-laki dari *UCI Repository*, serta adanya perbandingan beberapa algoritma *machine learning* yang digunakan, yaitu *Decision Tree*, *Naïve Bayes*, *K-Nearest Neighbors*, *Decision Stupm*, dan *Random Forest*.

Penelitian (Adi & Wintarti, 2022) tentang Komparasi Metode *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), dan *Random Forest* untuk Prediksi Penyakit Gagal Jantung. Dalam penelitiannya ini membandingkan tiga metode yaitu Metode *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), dan *Random Forest* untuk memprediksi penyakit gagal jantung. Data yang digunakan adalah *Heart Failure Clinical Records Dataset* dari *UCI Machine Learning Website*. Hasil penelitian menunjukkan akurasi metode SVM adalah



97%, metode *Random Forest* adalah 97%, dan metode KNN dengan parameter  $n\_neighbors = 20$  adalah 93%. Penelitian ini juga menciptakan program Klasifikasi Gagal Jantung dengan *Graphical User Interface* (GUI). Perbedaan penelitian tersebut dengan penelitian yang dilakukan oleh peneliti terdapat pada data yang digunakan yaitu *Heart Failure Clinical Record Dataset* dari *UCI Machine Learning*, serta adanya perbandingan beberapa algoritma *machine learning* yang digunakan, yaitu *Support Vector Machine* (SVM), *K-Nearest Neighbors*, dan *Random Forest*.

Tabel 2. 1 Penelitian terdahulu

No	Referensi	Input	Output	Metode Penelitian	Hasil
1	(Qomariyah et al., 2018)	Data rekam medis pasien dengan kode diagnose "I" ( <i>suspect</i> penyakit jantung) dan pasien yang berobat ke poli jantung di RS Awal Bros Bekasi	Akurasi klasifikasi sebesar 96,47%.	<i>Artificial Neural Network</i>	Hasil penelitian ini menunjukkan akurasi klasifikasi yang didapat sebesar 96,47%.
2	(Sumwiza et al., 2023)	<i>Heart Disease Dataset</i> dengan jumlah 1025 dan 14 atribut	Akurasi klasifikasi yang didapat sebesar 99%	<i>Random Forest</i>	Hasil penelitian ini menunjukkan akurasi klasifikasi yang didapat sebesar 99%.
3	(Scarano et al., 2023)	Dataset yang terdiri dari 72.363 kecelakaan pengendara sepeda yang terjadi di Inggris dari tahun 2016 hingga tahun 2018	<i>Random Parameters Logit Model</i> menunjukkan heterogenies signifikan yang tidak teramati pada data dan memiliki kecocokan yang baik dengan nilai <i>Mcfadden Pseudo R2</i>	<i>Random Forest</i> dan <i>Random Parameters Logit Model</i> .	Hasil penelitian ini menunjukkan algoritma <i>Random Forest</i> menunjukkan kinerja prediktif yang unggul dalam <i>F-measure</i> dan <i>G-mean</i> yang menunjukkan keakuratannya dalam mengklasifikasikan kecelakaan pengendara sepeda yang fatal dan parah. <i>Random Parameters Logit Model</i>

No	Referensi	Input	Output	Metode Penelitian	Hasil
			sebesar 0,21.		menunjukkan heterogenies signifikan yang tidak teramati pada data dan memiliki kecocokan yang baik dengan nilai <i>Mcfadden Pseudo R2</i> sebesar 0,21.
4	(Depari et al., 2022)	<i>Personal Key Indicators of Heart Disease Dataset</i>	Akurasi metode <i>Decision Tree</i> adalah 71%, <i>Naïve Bayes</i> adalah 72%, dan <i>Random Forest</i> adalah 75%	<i>Decision Tree</i> , <i>Naïve Bayes</i> , dan <i>Random Forest</i>	Hasil penelitian menunjukkan akurasi metode <i>Decision Tree</i> adalah 71%, <i>Naïve Bayes</i> adalah 72%, dan <i>Random Forest</i> adalah 75%.
5	(Nugraha, 2021)	<i>Heart Failure Prediction Dataset</i>	Akurasi <i>Random Forest</i> adalah 77%, <i>SVM</i> adalah 80%, <i>Gradient Boosting Machine</i> adalah 78%, <i>XGBOOST</i> adalah 80%, dan <i>Light GBM</i> adalah 68%	<i>Random Forest</i> , <i>Support Vector Machine (SVM)</i> , <i>Gradient Boosting Machine</i> , <i>XGBOOST</i> , dan <i>Light GBM</i>	Hasil penelitian ini menunjukkan akurasi <i>Random Forest</i> adalah 77%, <i>SVM</i> adalah 80%, <i>Gradient Boosting Machine</i> adalah 78%, <i>XGBOOST</i> adalah 80%, dan <i>Light GBM</i> adalah 68%.
6	(Annisa, 2019)	Data publik penderita penyakit jantung laki-laki dari <i>UCI Repository</i>	Akurasi <i>Decision Tree</i> adalah 76%, <i>Naïve Bayes</i> adalah 61%, <i>k-Nearest Neighbour</i> adalah 79%, <i>Random Forest</i> adalah 80%, dan <i>Decision Stump</i> adalah 79%	<i>Decision Tree</i> , <i>Naïve Bayes</i> , <i>k-Nearest Neighbour</i> , <i>Random Forest</i> , dan <i>Decision Stump</i>	Hasil penelitian ini menunjukkan akurasi <i>Decision Tree</i> adalah 76%, <i>Naïve Bayes</i> adalah 61%, <i>k-Nearest Neighbour</i> adalah 79%, <i>Random Forest</i> adalah 80%, dan <i>Decision Stump</i> adalah 79%.
7	(Adi & Wintarti, 2022)	<i>Heart Failure Clinical Records Dataset</i> dari <i>UCI Machine Learning Website</i>	akurasi metode <i>SVM</i> adalah 97%, metode <i>Random Forest</i> adalah 97%, dan metode <i>KNN</i> dengan parameter $n\_neighbors = 20$ adalah 93%, <i>GUI</i> .	Metode <i>Support Vector Machine (SVM)</i> , <i>K-Nearest Neighbors (KNN)</i> , dan <i>Random Forest</i>	. Hasil penelitian menunjukkan akurasi metode <i>SVM</i> adalah 97%, metode <i>Random Forest</i> adalah 97%, dan metode <i>KNN</i> dengan parameter $n\_neighbors = 20$ adalah 93%. Penelitian ini juga menciptakan program <i>Klasifikasi Gagal Jantung</i>

No	Referensi	Input	Output	Metode Penelitian	Hasil
					dengan <i>Graphical User Interface</i> .

Berdasarkan beberapa penelitian diatas, metode *Random Forest* mendapatkan akurasi yang baik dalam klasifikasi. Maka dari itu, penelitian ini dilakukan untuk mengetahui performa metode *Random Forest* dalam membangun model klasifikasi risiko serangan jantung dengan menggunakan bahasa pemrograman *Python3* dan *dataset* yang didapat dari *Kaggle* <https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset> yaitu *Heart Attack Analysis & Prediction Dataset*. Pada penelitian ini akan digunakan 13 atribut dari total 14 atribut pada data *Heart Attack Analysis & Prediction Dataset* (Rohayati & Widani, 2020).

Novelty atau pembaruan pada penelitian ini adalah dilakukannya proses *Synthetic Minority Oversampling Technique (SMOTE)* yang tidak dilakukan pada penelitian Sumwiza (2023). Adanya parameter *n\_estimator*, *max\_depth*, dan *bootstrap*. Selain itu, terdapat tiga model yaitu model A dengan pembagian data latih (*train*) sebesar 90%, dan data uji (*test*) sebesar 10%. Model B dengan data latih (*train*) sebesar 80%, dan data uji (*test*) sebesar 20%. Model C dengan data latih (*train*) sebesar 75%, dan data uji (*test*) sebesar 25% seperti pada penelitian (Sumwiza et al., 2023).

## 2.2 Serangan Jantung

Serangan jantung, atau yang dikenal dalam medis sebagai infark miokard, merupakan suatu kondisi dimana aliran darah ke jantung sangat berkurang atau

tersumbat dikarenakan pembuluh arteri mengalami penyempitan dikarenakan adanya sumbatan sehingga mengakibatkan suplai dan kebutuhan oksigen dalam darah tidak sesuai (Amrullah et al., 2022).

Penyumbatan terjadi Ketika ada penumpukan timbunan lemak pada dinding arteri yang memasok darah ke jantung sehingga menyebabkan berkurangnya pasokan oksigen ke otot jantung. Jika aliran darah tidak segera pulih, jaringan di jantung yang terkena akan mati (Mount Elizabeth Hospital, n.d.).

يَا أَيُّهَا النَّاسُ كُلُوا مِمَّا فِي الْأَرْضِ حَلَالًا طَيِّبًا وَلَا تَتَّبِعُوا خُطُوَاتِ الشَّيْطَانِ ، إِنَّهُ لَكُمْ عَدُوٌّ مُّبِينٌ  
ثُمَّ لَقَطَعْنَا مِنْهُ الْوَتِينَ

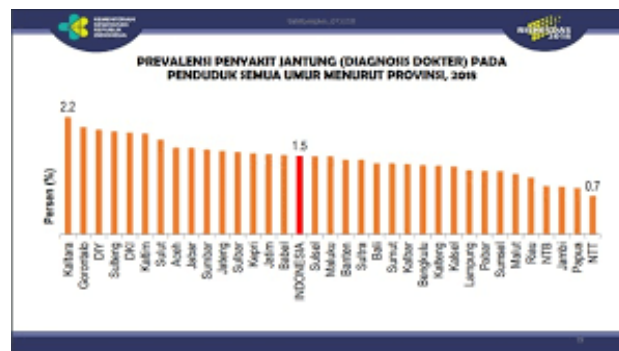
*“Wahai manusia, makanlah yang halal dan baik dari apa yang terdapat di bumi dan janganlah kamu mengikuti langkah-langkah syaithan; karena sesungguhnya syaithan adalah musuh yang nyata bagimu” (Q.S Al-Baqarah: 168).*

Menurut penjelasan Ibnu Katsir mengenai ayat ini, Allah telah mengizinkan manusia untuk memakan semua makanan yang ada di bumi yang halal, baik, bermanfaat dan tidak berbahaya bagi jiwa dan raga.

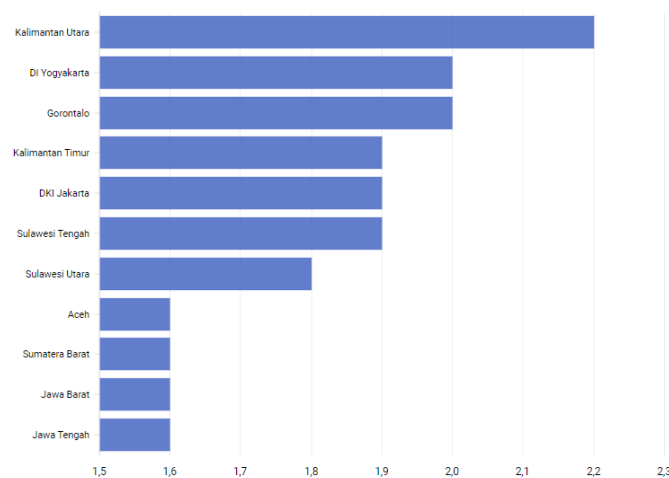
Penyebab paling umum dari serangan jantung adalah penyempitan satu atau lebih arteri yang memasok darah ke jantung. Kondisi ini disebabkan oleh penumpukan lemak atau kolesterol pada dinding arteri yang secara ilmiah disebut sebagai aterosklerosis. Akibatnya, aliran darah ke otot jantung menjadi tidak lancar sehingga suplai oksigen terganggu. Jika penyumbatan terjadi secara tiba-tiba dan parah, kondisi tersebut dapat menyebabkan serangan jantung. Faktor-faktor seperti merokok, *hipertensi*, *diabetes mellitus*, obesitas, faktor psikososial (depresi, stress, tekanan keuangan, konflik keluarga), kurangnya konsumsi buah

atau sayur, konsumsi alkohol, kurangnya aktifitas gerak dinilai dapat memicu terjadinya serangan jantung. Peningkatan risiko yang terkait dengan diabetes, alkohol, dan kurangnya olahraga dan *hipertensi* ditemukan lebih tinggi pada wanita. Sedangkan pada laki-laki cenderung terkena risiko serangan jantung pada usia dini (Niranjan & Dhamoon, 2023).

Data Riskesdas tahun 2018 menunjukkan prevalensi penyakit jantung berdasarkan diagnosis medis di Indonesia sebesar 1,5% dengan prevalensi tertinggi di provinsi Kalimantan Utara sebesar 2,2% (Kementerian Kesehatan, 2021).



Gambar 2. 1 Prevalensi Penyakit Jantung Tahun 2018



Gambar 2. 2 Grafik Penyakit Jantung di Indonesia

### 2.3 *Exploratory Data Analysis*

*Exploratory Data Analysis* (EDA) dalam bahasa Indonesia disebut sebagai Analisis Data Eksploratif atau Penganalisisan Data Eksploratif adalah proses menganalisis dan memvisualkan data untuk mendapatkan wawasan dan memahami pola dan hubungan yang mendasarinya. Hal ini melibatkan pemeriksaan data dari berbagai sudut pandang, mengidentifikasi potensi *outlier* atau nilai yang hilang, dan meringkas karakteristik utamanya. EDA membantu para peneliti dan analis untuk memahami data dengan lebih baik sebelum menerapkan teknik statistik atau *machine learning* (Sahoo et al., 2019). Sebagaimana telah difirmankan oleh Allah pada Al-Qur'an Surah Shad ayat 29.

كُتِبَ أَنْزَلْنَاهُ إِلَيْكَ مُبْرَكًا لِيَذَّبَرُواْ عَائِبَهُ وَلِيَتَذَكَّرَ أُولُوْاْ الْأَلْبَابِ

*“Ini adalah sebuah kitab yang Kami turunkan kepadamu penuh dengan berkah supaya mereka memperhatikan ayat-ayat-Nya dan supaya mendapat pelajaran orang-orang yang mempunyai pikiran” (Q.S. Shad: 29)*

Seperti yang sudah dijelaskan pada ayat diatas, kita diperintah untuk memperhatikan ayat-ayat-Nya. Hal tersebut merupakan sesuatu yang dilakukan pada EDA, dimana dilakukan pengekplorasi data agar lebih memahami tentang data tersebut. Sehingga dapat memanfaatkan data tersebut dengan tepat.

Menurut tafsir Ibnu Katsir surah Shad ayat 29 diatas menekankan pentingnya menelaah dan memahami Al-Qur'an yang mengharuskan untuk memahami maknanya, sama seperti EDA yang dilakukan untuk memahami karakteristik data yang akan dianalisa. Ayat tersebut juga mendorong kita untuk berpikir kritis terkait makna yang terkandung dalam Al-Qur'an. Demikian pula

dalam EDA yang melakukan identifikasi adanya *anomaly* atau *outliers* dalam data yang mungkin mempengaruhi hasil analisis selanjutnya.

Pada dasarnya, EDA adalah sebuah pendekatan untuk melihat apa yang dapat dikomunikasikan oleh data, terlepas dari pemodelan formal atau tugas pengujian hipotesis. EDA membantu menganalisis kumpulan data untuk meringkas karakteristik statistiknya yang berfokus pada empat aspek utama, seperti, ukuran tendensi sentral (rata-rata, modus, dan median), ukuran penyebaran (deviasi standar dan varian), bentuk distribusi dan keberadaan pencilan.

EDA biasanya merupakan langkah pertama dalam pengolahan data. EDA memungkinkan untuk memvisualisasikan data untuk memahaminya serta membuat hipotesis untuk analisis lebih lanjut. Analisis eksplorasi berpusat pada pembuatan sinopsis data atau wawasan untuk langkah selanjutnya dalam pengolahan data. EDA aktualnya mengungkapkan kebenaran dasar tentang data tanpa membuat asumsi yang mendasarinya. EDA ini biasa digunakan sebagai proses untuk lebih memahami jenis pemodelan dan hipotesis yang dapat dibuat. Komponen utama dari EDA seperti meringkas data, analisis statistik, dan visualisasi data (Mukhiya & Ahmed, 2020).

Eksplorasi data berfungsi sebagai langkah penting dalam proses analisis data yang menawarkan berbagai manfaat. Pada penelitian ini *Exploratory Data Analysis* atau EDA digunakan untuk visualisasi data, seperti keberadaan *outliers*, hubungan data dengan kelas target. Pada akhirnya proses ini membuka jalan bagi pemilihan teknik statistik atau *machine learning* yang sesuai untuk analisis

selanjutnya, membangun pondasi yang kuat untuk pengambilan keputusan berbasis data dan pemecahan masalah.

## 2.4 Klasifikasi

Klasifikasi adalah proses mengkategorikan objek atau data ke dalam kategori atau kelas tertentu berdasarkan atribut atau karakteristiknya (Mourya et al., 2018). Menurut etimologinya, “klasifikasi” berasal dari bahasa Inggris *classification* dan bahasa Prancis *classification* yang mengacu pada metode pengelompokan data secara sistematis atau menurut beberapa aturan yang telah diterapkan sebelumnya. Berdasarkan Kamus Besar Bahasa Indonesia, klasifikasi adalah pengaturan kelompok atau kelas secara sistematis menurut aturan atau standar yang telah ditetapkan.

Dalam Al-Qur’an klasifikasi dijelaskan pada Surah Al-An’am ayat 141:

وَهُوَ الَّذِي أَنْشَأَ جَنَّاتٍ مَّعْرُوشَاتٍ وَغَيْرِ مَّعْرُوشَاتٍ وَالنَّخْلَ وَالزَّرْعَ مُخْتَلِفًا أُكُلُهُ وَالزَّيْتُونَ وَالرُّمَّانَ مُتَشَابِهًا  
وَوَغَيْرَ مُتَشَابِهٍ كُلُوا مِنْ ثَمَرِهِ إِذَا أَثْمَرَ وَآتُوا حَقَّهُ وَلَا تُسْرِفُوا إِنَّهُ لَا يُحِبُّ الْمُسْرِفِينَ

*“Dan Dia-lah yang menjadikan tanaman-tanaman yang merambat dan yang tidak merambat, pohon kurma, tanaman beraneka ragam rasanya, zaitun dan delima yang serupa (bentuk dan warnanya) dan tidak serupa (rasanya). Makanlah buahnya apabila ia berbuah dan berikanlah haknya (zakatnya) pada waktu memetik hasilnya, tapi janganlah berlebih-lebihan. Sesungguhnya Allah tidak menyukai orang yang berlebih-lebihan” (Q.S. Al-An’am:141)*

Berdasarkan tafsir Ibnu Katsir surah Al-An’am ayat 141 ini menjelaskan tentang ciptaan Allah SWT yang beragam termasuk tanaman. Allah SWT menciptakan tanaman dengan berbagai jenis, rasa, dan bentuk. Ayat ini juga mengingatkan manusia untuk bersyukur atas nikmat Allah SWT dengan cara



memakan buah-buahan tersebut dan mengeluarkan zakatnya. Sama seperti Allah SWT yang menciptakan tanaman dengan berbagai jenis, rasa, dan bentuk, teknik klasifikasi dalam data *mining* digunakan untuk mengklasifikasikan data yang beragam berdasarkan pola dan kategorinya.

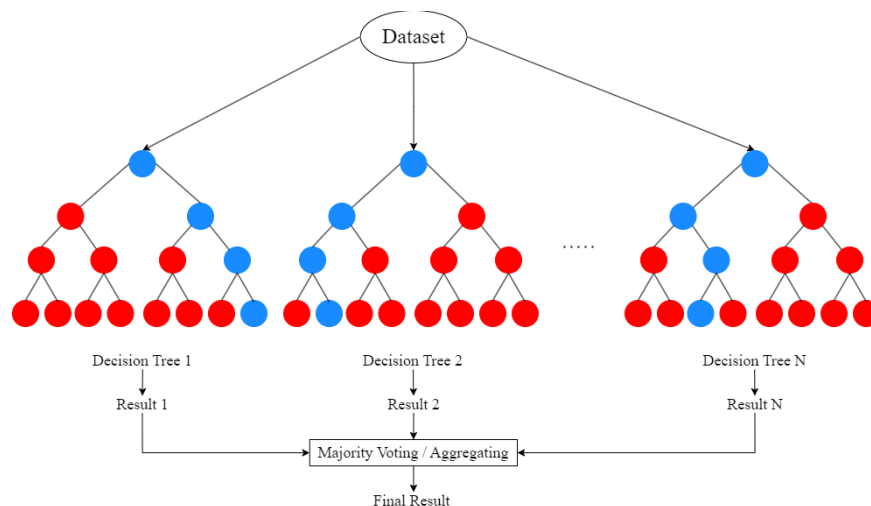
Metode klasifikasi digunakan untuk membagi item atau data ke dalam kelas atau kategori berbeda berdasarkan karakteristiknya. Dengan mengelompokkan objek yang serupa, tujuan utama klasifikasi adalah untuk mengatur dan memahami data dalam jumlah besar. Klasifikasi banyak digunakan di berbagai bidang termasuk *machine learning*, penggalian data, pencarian informasi, dan pengenalan gambar (Mourya et al., 2018).

Klasifikasi diperlukan dalam berbagai bidang dan aplikasi untuk mengatur dan mengkategorikan data atau dokumen dalam jumlah besar berdasarkan karakteristik atau atributnya. Klasifikasi memungkinkan pemrosesan dan analisis data yang efisien dengan mengelompokkan data yang serupa, sehingga dapat membantu dalam membuat keputusan yang tepat (Bhavani & Santhosh Kumar, 2021).

Dalam penelitian ini, algoritma klasifikasi digunakan untuk mengklasifikasikan kelas pasien yang kemungkinan memiliki risiko serangan jantung berdasarkan berbagai atribut datanya dengan kelas 0 (tidak berisiko serangan jantung) dan kelas 1 (berisiko serangan jantung) (Jindal et al., 2021). Algoritma *RandomForestClassifier* digunakan untuk mengklasifikasikan pasien ke dalam kategori yang berbeda, yaitu “risiko serangan jantung” atau “tidak ada risiko serangan jantung”.

## 2.5 Random Forest

*Random Forest* merupakan skema yang diusulkan oleh Leo Breiman pada tahun 2001 untuk membangun ansambel dengan sekumpulan pohon keputusan (*decision tree*) yang ada di subset data yang dipilih secara acak. Ini adalah algoritma yang menggabungkan beberapa pohon keputusan (*decision tree*) untuk membuat model (Jin et al., 2020). Setiap pohon keputusan dilatih pada subset data acak dan fitur yang pilih secara acak. Algoritma ini membuat klasifikasi dengan menggabungkan klasifikasi dari semua pohon keputusan dimana pohon yang dinilai terbaik oleh mesin yang akan digunakan.



Gambar 2. 3 Alur Metode *Random Forest*

Alur dasar dari algoritma *Random Forest Classifier* seperti pada gambar dimulai dengan data *training* dimuat ke dalam memori, kemudian dibagi menjadi subset acak menggunakan teknik *bootstrap*. Teknik *bootstrap* memungkinkan pengambilan sampel dengan penggantian dari data *training* asli, sehingga setiap subset bisa memiliki data yang berulang. Untuk setiap subset data yang dihasilkan, sebuah pohon keputusan dibuat. Pohon keputusan ini dibangun dengan

membatasi kedalaman pohon atau maksimum fitur yang dapat digunakan untuk setiap *split* untuk menghindari *overfitting*. Pada setiap *node* dalam pohon, *subset* fitur acak dipilih untuk mencari *split* terbaik, menambah variasi dan mengurangi korelasi antar pohon.

Proses pembuatan pohon keputusan ini diulang sebanyak  $n\_estimator$  kali, yang merupakan jumlah pohon yang diinginkan dalam model *ensemble*. Setelah seluruh pohon keputusan dibuat, model siap untuk melakukan prediksi terhadap data baru. Prediksi dilakukan dengan mengumpulkan dari setiap pohon keputusan. Untuk setiap pohon, data baru diklasifikasikan dan hasil prediksinya dicatat. Setelah semua hasil dari setiap pohon terkumpul, dilakukan agregasi prediksi dengan menghitung jumlah prediksi untuk setiap kelas dari semua pohon keputusan. Terakhir, data baru diklasifikasikan ke dalam kelas dengan jumlah prediksi terbanyak yang artinya kelas yang paling sering diprediksi oleh pohon-pohon keputusan akan menjadi kelas akhir untuk data baru, mengikuti prinsip mayoritas suara dalam *ensemble learning*.

أَمْ تَرَىٰ كَيْفَ ضَرَبَ اللَّهُ مَثَلًا كَلِمَةً طَيِّبَةً كَشَجَرَةٍ طَيِّبَةٍ أَصْلُهَا ثَابِتٌ وَفَرْعُهَا فِي السَّمَاءِ

“Tidakkah kamu perhatikan bagaimana Allah telah membuat perumpamaan kalimat yang baik seperti pohon yang baik, akarnya teguh dan cabangnya (menjulang) ke langit” (Q.S. Ibrahim: 25).

Sebagaimana diriwayatkan oleh As-Saddi, dari Murrah, dari Ibnu Ma’sud yang mengatakan pohon yang dimaksud adalah pohon kurma (Tafsir Ibnu Katsir, n.d.). Seperti halnya pohon yang baik atau dalam algoritma *Random Forest* yang merupakan sekumpulan pohon keputusan (*decision tree*) yang ada di subset data

yang dipilih secara acak dan terbaik menurut mesin sehingga diharapkan dapat menghasilkan yang terbaik.

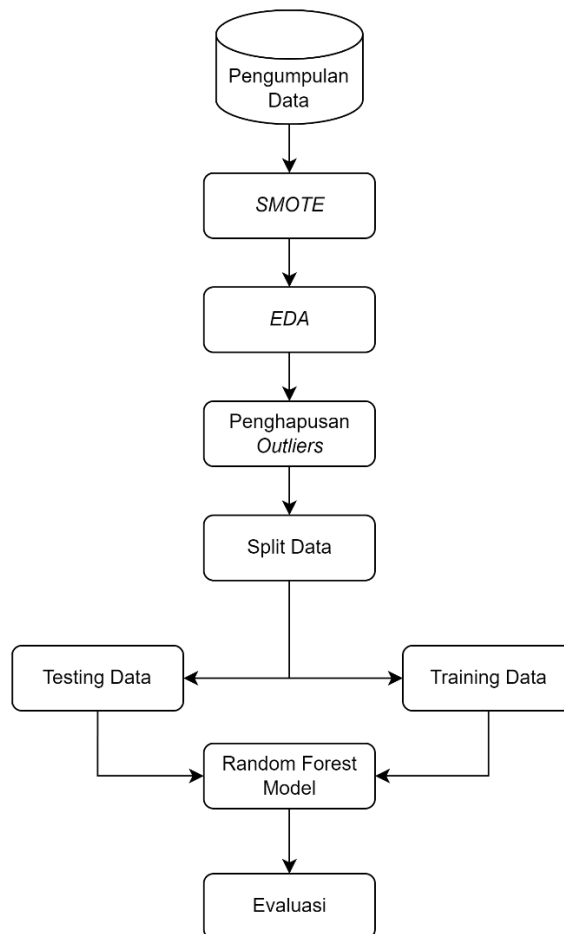
Menurut tafsir Ibnu Katsir surah Ibrahim ayat 25 menjelaskan tentang perumpamaan kalimat yang baik seperti pohon yang baik. Pohon yang baik memiliki akar yang teguh dan cabangnya yang menjulang ke langit. Keterkaitan antara ayat di atas dengan metode *Random Forest* yaitu sama seperti pohon yang baik yang memiliki banyak cabang yang kuat, *Random Forest* juga memiliki banyak pohon keputusan yang bekerja sama untuk menghasilkan hasil yang lebih baik.

Keuntungan utama dari *Random Forest* terletak pada kemampuannya untuk mengolah data berdimensi tinggi dan mengolah variabel yang mengandung *noise*. *Random Forest* juga menunjukkan konsistensi dan kemampuan untuk menyesuaikan dengan data yang jarang, yang berarti bahwa tingkat konvergensi algoritma hanya bergantung pada jumlah fitur penting yang ada, bukan pada seberapa banyak variabel *noise* yang ada dalam data (Klusowski, 2018). Selain itu *Random Forest* membuat kesalahan lebih sedikit, memberikan hasil klasifikasi yang baik seperti pada beberapa penelitian yang telah dipaparkan, dapat secara efektif menangani data latih yang sangat besar, dan digunakan untuk memperkirakan data yang hilang. *Random Forest* juga memuat matrik kesalahan, kekuatan, korelasi, dan kepentingan variabel, yang membantu dalam memahami dan menginterpretasikan kinerja model dan signifikansi fitur individu. Dalam penelitian ini metode *Random Forest* digunakan untuk mengklasifikasikan *dataset* berdasarkan kelas target yang telah ditentukan.

## BAB III METODOLOGI PENELITIAN

### 3.1 Desain Sistem

Desain sistem ini dimulai dengan pengumpulan data dan *preprocessing* untuk mengimplementasikan analisis data dan pengembangan model klasifikasi untuk klasifikasi risiko serangan jantung. Dalam proses membangun model untuk klasifikasi risiko serangan jantung dilakukan pengumpulan data, *SMOTE*, *EDA*, penghapusan *outliers*, *split data*, membangun model *Random Forest*, dan Evaluasi.



Gambar 3. 1 Desain Sistem

### 3.2 Pengumpulan Data

Pada tahapan ini akan dilakukan pengumpulan data menggunakan data sekunder. Data yang digunakan dalam penelitian ini yaitu dataset “*Heart Attack Analysis & Prediction Dataset*” yang merupakan *open access* data web dari Kaggle dengan alamat [kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset/data](https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset/data). Total data yang didapat berjumlah 303 data dengan 14 atribut seperti pada tabel penjelasan dibawah.

Tabel 3. 1 Detail Fitur Dataset

Fitur	Keterangan
<i>Age</i>	Usia pasien
<i>Sex</i>	Jenis kelamin pasien
<i>CP – chest pain</i>	Jenis nyeri dada yang diderita pasien. Dimana terdapat 4 nilai yaitu: Tipe 1: <i>typical angina</i> Tipe 2: <i>atypical angina</i> Tipe 3: <i>non-anginal pain</i> Tipe 4: <i>asymptomatic</i>
<i>Trestbps – resting blood pressure</i>	Tekanan darah pasien Ketika dalam keadaan istirahat menggunakan satuan mg/dl
<i>Chol</i>	Kadar kolesterol dalam darah pasien menggunakan satuan mg/dl
<i>FBS – fasting blood sugar</i>	Kadar gula darah pasien, dengan 2 nilai, dimana 1 jika kadar gula darah pasien lebih dari 120 mg/dl, dan 0 jika gula darah pasien kurang dari atau sama dengan 120 mg/dl.
<i>Restecg</i>	Kondisi EKG pasien Ketika dalam keadaan istirahat dengan 3 nilai, dimana 0 untuk keadaan normal, 1 untuk <i>ST-T wave abnormality</i> yaitu keadaan gelombang inversions T atau ST meningkat maupun menurun lebih dari 0,5 mV, dan nilai 2 untuk keadaan ventricular kiri mengalami hipertropi.
<i>Thalach</i>	Rata-rata detak jantung pasien.
<i>Exang</i>	Keadaan dimana pasien akan mengalami nyeri dada apabila berolahraga, terdapat 2 nilai, dimana 0 tidak nyeri, dan 1 apabila menyebabkan nyeri.
<i>Oldpeak</i>	Penurunan ST akibat olahraga
<i>Slope</i>	<i>Slope</i> dari puncak ST setelah berolahraga dengan 3 nilai, dimana 0 untuk <i>downsloping</i> , 1 untuk <i>flat</i> , dan 2 untuk <i>upsloping</i> .
<i>Ca</i>	Banyaknya pembuluh darah yang terdeteksi melalui proses pewarnaan flourosopy
<i>Thal</i>	Detak jantung pasien dengan 3 nilai, dimana 1 berarti <i>fixed defect</i> , 2 berarti normal, dan 3 yang berarti <i>reversible defect</i>
<i>Num</i>	Hasil diagnosa penyakit jantung dengan 2 nilai, dimana 0 jika penyempitan diameter kurang dari atau sama dengan 50%, dan 1 jika

<b>Fitur</b>	<b>Keterangan</b>
	penyempitan diameter lebih dari 50%.

Tabel 3. 2 Sample Data

<i>Age</i>	<i>Sex</i>	<i>Chest pain</i>	<i>Trtbps</i>	<i>Chol</i>	<i>FBS</i>	<i>Restecg</i>	<i>Thalachh</i>	<i>exng</i>	<i>Oldspeak</i>	<i>Slope</i>	<i>Caa</i>	<i>Thall</i>	<i>Num (output)</i>
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
44	1	1	120	263	0	1	173	0	0	2	0	3	1
52	1	2	172	199	1	1	162	0	0.5	2	0	3	1



### 3.3 *Preprocessing*

#### 3.3.1 Pemisahan Data

Tahapan awal *preprocessing* data dalam penelitian ini adalah memisahkan data menjadi fitur ( $x$ ) dan target ( $y$ ). Proses ini penting dilakukan pada tahap *preprocessing* data sebelum menerapkan model *machine learning*. Pendekatan ini memungkinkan model untuk memahami pola data dan mengklasifikasi variabel target ( $y$ ) berdasarkan fitur ( $x$ ).

#### 3.3.2 *Synthetic Minority Oversampling Technique (SMOTE)*

Pada tahapan *preprocessing* data, dilakukan *Synthetic Minority Oversampling Technique (SMOTE)* untuk menyeimbangkan *imbalanced* data atau data yang tidak seimbang. Terdapat *imbalanced* data pada *dataset* yang digunakan untuk penelitian ini, sehingga diperlukan Teknik *SMOTE* untuk menanganinya sehingga distribusinya seimbang. Kelas 0 (tidak berisiko serangan jantung) memiliki 138 data, sedangkan kelas 1 (berisiko serangan jantung) memiliki 165 data. Meskipun memiliki perbedaan yang tidak begitu signifikan, namun peneliti tetap melakukan teknik *SMOTE* untuk menangani ketidakseimbangan data pada *dataset* dengan tujuan agar dapat menghasilkan klasifikasi yang lebih baik.

Tahapan melakukan *Synthetic Minority Oversampling Technique (SMOTE)* yaitu:

- a. Memilih satu sampel data minoritas secara acak
- b. Menentukan tetangga terdekat ( $k$ ) dari sampel minoritas yang dipilih

- c. Buat data sintesis baru dengan menambahkan kombinasi terbobot antara sampel minoritas yang dipilih oleh salah satu tetangga terdekatnya, dengan menggunakan persamaan berikut:

$$x_{syn} = x_i(x_{knn} - x_i) \times \sigma \quad (3.1)$$

Contoh perhitungannya sebagai berikut:

Misalkan terdapat dataset dengan 2 kelas: mayoritas (kelas 0) dan minoritas (kelas 1). Distribusi kelasnya 80% (mayoritas) dan 20% (minoritas). Langkah-langkahnya yaitu:

- a. Sampel yang dipilih adalah [1,2].
- b.  $k = 3$ , maka  $k = [0,3], [2,1], [1,3]$
- c. Buat data sintesis baru dengan menambahkan kombinasi terbobot antara sampel minoritas yang dipilih oleh salah satu tetangga terdekatnya, dengan menggunakan persamaan berikut:

$$\begin{aligned} x_{syn} &= [1,2] + (0.5 \\ &\quad \times ([0,3] - [1,2])) \\ x_{syn} &= [1.5, 2.5] \end{aligned}$$

### 3.1 *Exploratory Data Analysis (EDA)*

*Exploratory Data Analysis (EDA)* adalah tahap penting dalam analisis data, dengan tujuan untuk memahami karakteristik dataset, mengidentifikasi pola atau tren, dan memvisualisasikan informasi penting sebelum mengembangkan model atau membuat keputusan analisis lebih lanjut. Pada penelitian ini EDA difokuskan untuk memvisualisasikan distribusi variabel numerikal.

### 3.4.1 Visualisasi Distribusi Variabel Numerikal

Proses dari EDA melibatkan visualisasi distribusi variabel numerik menggunakan *boxplot*. Langkah awalnya adalah mengidentifikasi variabel numerik yang akan divisualisasikan, meliputi 'age', 'trtbps', 'chol', 'thalachh', dan 'oldpeak'. Variabel-variabel ini menandakan berbagai aspek klinis yang dapat berdampak pada risiko serangan jantung.

Tahapan pertama yang dilakukan adalah mengidentifikasi variabel numerik yang relevan untuk mendapat pemahaman yang lebih dalam mengenai karakteristik dan distribusi data. Melalui visualisasi ini, dapat diketahui pencilan atau *outliers*, rentang nilai, median, kuartil, serta distribusi nilai-nilai yang memungkinkan. Analisis diperluas dengan memvisualisasikan hubungan antara variabel numerikal terhadap kelas target dengan menggunakan *boxplot*. Analisis ini memberikan wawasan dalam memahami kontribusi variabel numerical terhadap kelas target yang dituju.

### 3.2 Penanganan *Outliers*

Proses penanganan *outliers* dilakukan setelah tahap eksplorasi data atau EDA. Dalam implementasinya dilakukan identifikasi indeks *outlier* berdasarkan kriteria yang telah ditentukan yaitu menggunakan *interquartile range* (IQR) dengan *kuartil* pertama atau Q1 merupakan 25% bagian terbawah data, *kuartil* ketiga atau Q3 merupakan 75% bagian teratas data, dan batas *outlier* yaitu 1,5 kali IQR yang digunakan untuk menentukan batas atas dan batas bawah untuk mendeteksi *outliers*. Setelah *outliers* teridentifikasi, *dataset* diperbarui dengan menghapus baris yang mengandung *outlier*.

Tahapan untuk melakukan penghapusan *outliers* adalah sebagai berikut:

- a. Menghitung  $Q_1$  dan  $Q_3$
- b. Menentukan *Interquartile Range* (IQR), dengan menggunakan persamaan berikut:

$$IQR = Q_3 - Q_1 \quad (3.2)$$

Keterangan:

$Q_3$ , quartil ketiga: nilai yang membagi data menjadi 75% bagian terbawah dan 25% bagian teratas.

$Q_1$ , quartil pertama: nilai yang membagi data menjadi 25% bagian terbawah dan 75% bagian teratas.

- c. Menentukan batas *outliers* berdasarkan IQR dengan menggunakan persamaan berikut:

$$\begin{aligned} \text{Batas bawah} &= Q_1 \\ &\quad - 1,5 \times IQR \end{aligned} \quad (3.3)$$

$$\text{Batas atas} = Q_3 - 1,5 \times IQR \quad (3.4)$$

- d. Mengidentifikasi *outliers*
- e. Menghapus baris yang terdapat *outliers*.

Contoh perhitungan untuk melakukan penghapusan *outliers* adalah sebagai berikut:

Misalkan terdapat dataset sebagai berikut:

[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]

- a. Menghitung Q1 dan Q3

$$Q1 = \text{Median}(X_{\{(1:((n+1)/4))\}}) \quad (3.5)$$

$$Q1 = 5.5$$

$$Q3 = \text{Median}(X_{\{((n+3)/4):n\}}) \quad (3.6)$$

$$Q3 = 12.5$$

- b. Menghitung IQR

$$IQR = Q_3 - Q_1$$

$$IQR = 12.5 - 5.5$$

$$IQR = 7$$

- c. Menghitung batas atas dan batas bawah

$$\text{Batas bawah} = Q_1 - 1,5 \times IQR$$

$$\text{Batas bawah} = 5.5 - 1,5 \times 7$$

$$\text{Batas bawah} = -2$$

$$\text{Batas atas} = Q_3 - 1,5 \times IQR$$

$$\text{Batas atas} = 12.5 - 1,5 \times 7$$

$$\text{Batas atas} = 23$$

- d. Identifikasi *outliers*

Data yang bernilai lebih kecil dari -2 atau lebih besar dari 23 dianggap sebagai *outliers*

- e. Menghapus *outliers*

Jika terdapat *outliers* maka dilakukan penghapusan pada baris yang mengandung *outliers*

### 3.3 Implementasi Metode *Random Forest*

Setelah data di-*split* yaitu data latih (*train*), dan data uji (*test*). Beberapa *hyperparameter* dalam algoritma *Random Forest* digunakan dalam penelitian ini. Mengatur jumlah maksimal pohon keputusan yang akan dibangun dalam *ensemble* dengan menggunakan *n\_estimator*. Nilai *seed* atau nilai acak dalam algoritma dengan menggunakan *random\_state* untuk mengatur inisialisasi model yang melibatkan unsur keacakan dan hasil yang dapat direproduksi. Kedalaman maksimum dari setiap pohon juga diatur dengan menggunakan *max\_depth*. Serta, pengambilan sampel dengan penggantian dilakukan saat membangun setiap pohon dalam *ensemble Random Forest* dengan menggunakan *bootstrap = true*.

Tahapan pelatihan model dengan data latih (*train*) yang digunakan untuk melatih model dibagi menjadi dua, yaitu fitur data latih dan target klasifikasi yang sesuai. Setelah data latih dibagi, dilakukan pembagian proporsi data pengujian untuk tahapan pelatihan model. Setelah pelatihan selesai, model digunakan untuk membuat klasifikasi pada data uji. Hasil klasifikasi kemudian dievaluasi lebih lanjut menggunakan menggunakan beberapa matrik evaluasi.

Contoh perhitungan untuk membuat pohon keputusan dalam *Random Forest* adalah sebagai berikut:

- a. Misalkan terdapat dataset dengan dua fitur (X1 dan X2) dan satu target variabel (kelas). Dataset ini diklasifikasikan menjadi dua kelas: “Ya” dan “Tidak”.

Data = [[1,2], [3,4], [5,6], [7,8], [9, 10]]

Label = [“Ya”, “Tidak”, “Ya”, “Ya”, “Tidak”]

b. Membagi data (Bootsrap)

Asumsikan  $n_{estimator} = 2$

Subset 1:

$X_1 = [[3, 4], [5, 6], [9, 10]]$

$y_1 = ["Tidak", "Ya", "Tidak"]$

Subset 2:

$X_2 = [[1, 2], [7, 8]]$

$y_2 = ["Ya", "Ya"]$

c. Membangun pohon keputusan pada setiap subset dengan menggunakan persamaan

$$h_j(x) = \operatorname{argmax}_y P(y | x, T_j) \quad (3.7)$$

Keterangan:

$h_j(x)$  Q1: prediksi pohon keputusan ke-j untuk data x

y : label kelas

$P(y | x, T_j)$  probabilitas kelas y untuk data x pada pohon keputusan ke-j

$T_j$  adalah pohon keputusan ke-j

Rumus untuk voting:

$$\operatorname{pred}(x) = \operatorname{argmax}_y \operatorname{count}(h_j(x) = y) \quad (3.8)$$

Keterangan:

$\operatorname{pred}(x)$  Q1: prediksi akhir untuk data x

y: label kelas

$\operatorname{count}(h_j(x) = y)$ : jumlah pohon keputusan yang memprediksi kelas y untuk data x.

Pohon keputusan 1:

$X_1 \leq 4$ : Ya

$X_1 > 4$ :

$X_2 \leq 5$ : Tidak

$X_2 > 5$ : Ya

Pohon keputusan 2:

$X_1 \leq 2$ : Ya

$X_1 > 2$ :

$X_2 \leq 6$ : Ya

$X_2 > 6$ : Tidak

Misalkan akan memprediksi kelas untuk data baru: [11,12].

1. Hitung prediksi dengan setiap pohon keputusan:

Pohon keputusan 1:

$X_1 = 11 > 4$ , lanjut ke kanan

$X_2 = 12 > 5$ , prediksi “Tidak”.

Pohon keputusan 2:

$X_1 = 11 > 2$ , lanjut ke kanan

$X_2 = 12 > 6$ , prediksi “Tidak”.

2. Agregasi prediksi

Prediksi “Tidak” dari kedua pohon keputusan.

3. Klasifikasi data baru ke kelas “Tidak”.

### 3.4 Evaluasi

Hasil klasifikasi yang didapat dari pelatihan model dievaluasi lebih lanjut menggunakan beberapa matrik. Yang pertama adalah *Confusion Matrix*, yang memberikan gambaran tentang seberapa baik model dapat memprediksi *instance*



data dalam kategori benar atau salah. Proses ini melibatkan perhitungan *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dimana *True Positive* (TP) dan *True Negative* (TN) menggambarkan keakuratan klasifikasi yang sesuai dengan hasil yang aktualnya. Sedangkan *False Positive* (FP) dan *False Negative* (FN) terjadi ketika hasil yang diklasifikasi tidak sesuai dengan kondisi aktualnya (Erdiansyah et al., 2022).

Tabel 3. 3 *Confusion Matrix*

<i>Actual Class</i>	<i>Assigned Class</i>	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	<i>True Positive</i>	<i>False Negative</i>
<i>Negative</i>	<i>False Positive</i>	<i>True Negative</i>

Selanjutnya, akurasi model dihitung dan ditampilkan. Akurasi dilakukan untuk memberikan indikasi seberapa baik model dapat memberikan klasifikasi yang akurat. Nilai akurasi dapat dihitung menggunakan persamaan berikut:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.7)$$

Matrik selanjutnya adalah presisi yang menunjukkan seberapa baik model dalam mengidentifikasi *instance* positif. Nilai presisi dapat dihitung menggunakan persamaan berikut:

$$Precision = \frac{TP}{TP + FP} \quad (3.2)$$

Kemudian *recall* yang dihitung untuk menunjukkan seberapa baik model dalam mengidentifikasi semua *instance* positif. Nilai *recall* dapat dihitung menggunakan persamaan berikut:

$$Recall = \frac{TP}{TP + FN} \quad (3.3)$$

Selanjutnya, *F1 score* yang merupakan matrik gabungan yang memperhitungkan keseimbangan antara presisi dan *recall*. Nilai *F1 score* dapat dihitung menggunakan persamaan berikut:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3.4)$$

Tahapan terakhir adalah laporan klasifikasi yang menampilkan sejumlah matrik evaluasi untuk setiap kelas target, termasuk presisi, *recall*, dan *F1 score*.

### 3.5 Skenario Pengujian

Pada pemilihan model, metode *Random Forest* diimplementasikan menggunakan *RandomForestClassifier* dengan menggunakan parameters  $n\_estimator = 100$  yang artinya menggunakan 100 pohon keputusan dalam *ensemble (forest)*. Selain itu, terdapat juga parameter  $random\_state = 57$  yang merupakan *seed* untuk mengontrol *random number generator*. Parameter  $max\_depth = 50$  juga digunakan dalam pengujian. *Parameter* ini dipilih untuk digunakan pada penelitian ini setelah dilakukan pengujian sampel data dengan menggunakan beberapa kombinasi *parameter* seperti pada tabel berikut.

Tabel 3. 4 Tuning Hyperparameter

<b>Tuning Hyperparameter</b>	<b>Hasil Akurasi</b>
$n\_estimator = 100, random\_state = 57,$ $max\_depth = 50$	0.77
$n\_estimator = 25, random\_state = 42,$ $max\_depth = 35$	0.71
$n\_estimator = 50, random\_state = 27,$ $max\_depth = 25$	0.74

Pada tahapan pelatihan model data dibagi menjadi 3 model yang meliputi dua bagian utama, yaitu data latih (*train*), dan data uji (*test*). Peneliti membagi data untuk Model A dengan data *train* sebesar 90%, dan data *test* sebesar 10%. Model B dengan data *train* sebesar 80%, dan data *test* sebesar 20%. Model C dengan data *train* sebesar 75%, dan data *test* sebesar 25% (Sumwiza et al., 2023). Pembagian data ini membantu dalam pelatihan model dengan menggunakan data latih (*train*) dan menguji performa model menggunakan data uji (*test*). Data uji (*test*) digunakan untuk menguji performa model dengan matrik evaluasi menggunakan akurasi, presisi, recall, dan F1-score.

Pendekatan kasus uji akan dilakukan untuk membandingkan beberapa teknik pemrosesan data. Pertama, dengan membandingkan hasil dari data yang menggunakan metode *SMOTE* (*Syntetic Minority Over-sampling Technique*) dalam mengatasi *imbalanced* data dengan situasi tanpa penggunaannya. Penanganan *outlier* pada *dataset* juga akan dievaluasi, dengan satu kondisi melibatkan pemrosesan *outlier* dan kondisi lainnya tanpa penanganan *outlier*. Selain itu digunakan juga kasus uji dengan pembagian 3 model dan 4 kelompok data berdasarkan kombinasi fitur yang digunakan.

Tabel 3. 5 Skenario Pengujian

No	Kasus Uji		Pengujian
	Fitur	Model	
1	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol</i>	Model A dengan perbandingan data 90:10	Dengan <i>SMOTE</i> Selanjutnya Penghapusan <i>Outliers</i>
2	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps</i>		
3	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps, Slp, Sex</i>		
4	Semua Fitur		
5	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol</i>	Model B dengan perbandingan data 80:20	
6	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps</i>		
7	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps, Slp, Sex</i>		
8	Semua Fitur		
9	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol</i>	Model C dengan perbandingan data 75:25	
10	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps</i>		
11	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps, Slp, Sex</i>		
12	Semua Fitur		
13	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol</i>	<i>Data Real</i>	
14	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps</i>		
15	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps, Slp, Sex</i>		
16	Semua Fitur		
17	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol</i>	Model A dengan perbandingan data 90:10	

No	Kasus Uji		Pengujian	
	Fitur	Model		
18	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps</i>			
19	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps, Slp, Sex</i>			
20	Semua Fitur			
21	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol</i>	Model B dengan perbandingan data 80:20		
22	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps</i>			
23	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps, Slp, Sex</i>			
24	Semua Fitur			
25	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol</i>	Model C dengan perbandingan data 75:25		
26	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps</i>			
27	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps, Slp, Sex</i>			
28	Semua Fitur			
29	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol</i>	<i>Data Real</i>		
30	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps</i>			
31	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps, Slp, Sex</i>			
32	Semua Fitur			
33	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol</i>	Model A dengan perbandingan data 90:10		Dengan <i>SMOTE</i> dan Tanpa Penghapusan <i>Outliers</i>
34	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps</i>			
35	<i>Oldpeak, Thalachh, CP, Caa, Thall,</i>			

No	Kasus Uji		Pengujian
	Fitur	Model	
	<i>Age, Chol, Exng, Trtbps, Slp, Sex</i>		Dengan Penghapusan <i>Outliers</i> dan Tanpa <i>SMOTE</i>
36	Semua Fitur		
37	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol</i>	Model B dengan perbandingan data 80:20	
38	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps</i>		
39	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps, Slp, Sex</i>		
40	Semua Fitur		
41	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol</i>	Model C dengan perbandingan data 75:25	
42	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps</i>		
43	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps, Slp, Sex</i>		
44	Semua Fitur		
45	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol</i>	<i>Data Real</i>	
46	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps</i>		
47	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps, Slp, Sex</i>		
48	Semua Fitur		
49	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol</i>	Model A dengan perbandingan data 90:10	
50	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps</i>		
51	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps, Slp, Sex</i>		
52	Semua Fitur		
53	<i>Oldpeak, Thalachh, CP, Caa, Thall,</i>	Model B dengan perbandingan data	

No	Kasus Uji		Pengujian
	Fitur	Model	
	<i>Age, Chol</i>	80:20	
54	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps</i>		
55	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps, Slp, Sex</i>		
56	Semua Fitur		
57	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol</i>	Model C dengan perbandingan data 75:25	
58	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps</i>		
59	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps, Slp, Sex</i>		
60	Semua Fitur		
61	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol</i>	<i>Data Real</i>	
62	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps</i>		
63	<i>Oldpeak, Thalachh, CP, Caa, Thall, Age, Chol, Exng, Trtbps, Slp, Sex</i>		
64	Semua Fitur		

Tahapan terakhir dalam penelitian ini adalah evaluasi terhadap hasil kinerja model dengan menggunakan algoritma *Random Forest* dalam melakukan klasifikasi risiko serangan jantung. Evaluasi dilakukan dengan menggunakan beberapa matrik, yang pertama *Confusion Matrix* yang memberikan gambaran tentang seberapa baik model dapat memprediksi *instance* data dalam kategori benar atau salah pada setiap kelas. Selanjutnya beberapa matrik lain juga digunakan untuk mengukur kinerja model berdasarkan hasil *Confusion Matrix*. Matrik-matrik tersebut adalah *accuracy*, *recall*, *precision*, dan *f1-score*.

Secara keseluruhan, pada penelitian ini terdapat empat tahap skenario pengujian, yaitu:

- a. Pengujian terhadap *software Google Colab* yang digunakan.
- b. Pengujian dengan *tuning parameter* yaitu  $n\_estimator = 100$ ,  $max\_depth = 50$ , dan  $random\_state = 57$ .
- c. Skenario kasus uji dengan menggunakan 3 model, yang pada masing-masing model dilakukan dengan menggunakan 4 kelompok kombinasi fitur, dengan 4 pengujian yaitu menggunakan *SMOTE* selanjutnya dengan penghapusan *outliers*, menggunakan penghapusan *outliers* selanjutnya *SMOTE*, menggunakan *SMOTE* selanjutnya tanpa penghapusan *outliers*, dan menggunakan penghapusan *outliers* tanpa *SMOTE*.
- d. Pengujian keseluruhan data dengan menggunakan data yang didapat dari dataset "*Heart Attack Analysis & Prediction Dataset*" yang merupakan *open access* data web dari Kaggle.



## **BAB IV**

### **HASIL DAN PEMBAHASAN**

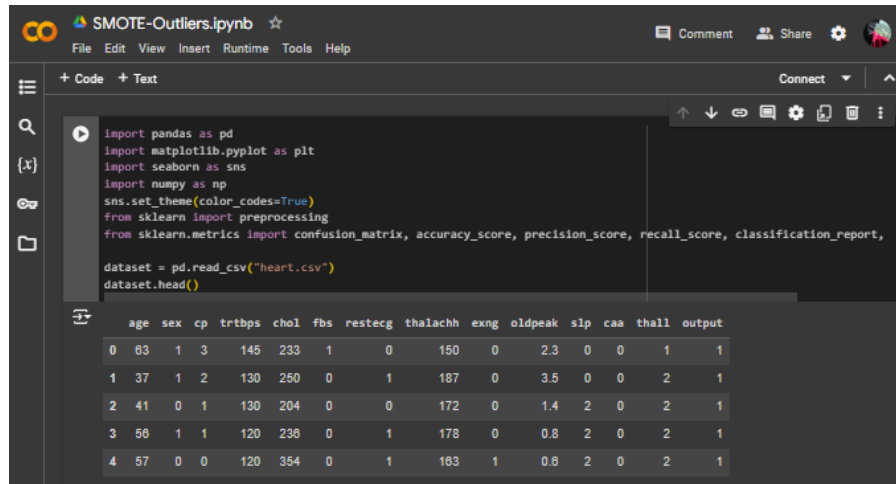
#### **4.1 Hasil Uji Coba**

Poin ini menjabarkan analisis hasil pengujian berdasarkan skenario pengujian sesuai pada sub bab 3.8 untuk mengetahui performa metode *Random Forest* dalam mengklasifikasikan risiko serangan jantung dengan menggunakan *dataset* yang berjumlah 303. Setelah dilakukan pengujian didapatkan beberapa hasil klasifikasi dan evaluasi matriknya dengan menggunakan *Random Forest*. Selanjutnya, sistem yang telah dibuat dievaluasi untuk melihat kinerja metode *Random Forest* yang mengklasifikasikan risiko serangan jantung dengan menggunakan beberapa matrik evaluasi.

##### **4.1.1 Pengujian *Software* yang Digunakan**

Hasil pengujian terhadap *Google Colab* sebagai *text editor* untuk membuat program klasifikasi dengan menggunakan bahasa pemrograman *Python3* menunjukkan bahwa *platform* ini efektif dan efisien. *Google Colab* menyediakan fitur yang terintegrasi dengan baik, mendukung berbagai *library Python* secara default, dan memungkinkan pengguna untuk menjalankan kode secara interaktif. Fitur-fitur seperti *auto-completion*, *syntax highlighting*, dan integrasi dengan *Google Drive* meningkatkan produktivitas dalam menulis dan menyimpan kode. Meski demikian, *Google Colab* memerlukan koneksi internet yang stabil dan beberapa keterbatasan terkait akses ke file sistem lokal. Secara keseluruhan *Google Colab* dapat digunakan sebagai alat yang efisien dan efektif untuk

pengembangan program klasifikasi dengan menggunakan bahasa pemrograman *Python3*.



```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
sns.set_theme(color_codes=True)
from sklearn import preprocessing
from sklearn.metrics import confusion_matrix, accuracy_score, precision_score, recall_score, classification_report,

dataset = pd.read_csv("heart.csv")
dataset.head()

```

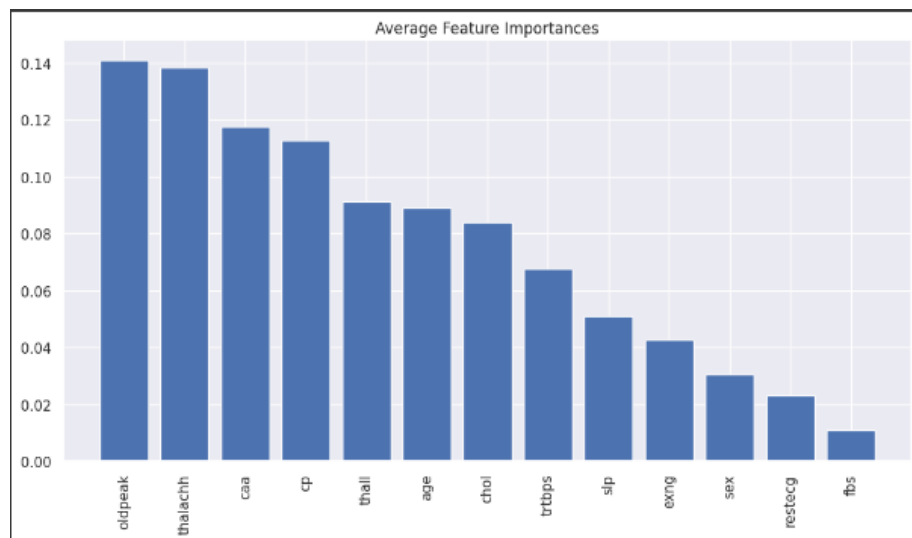
	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	238	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Gambar 4. 1 Hasil Pengujian *Software*

#### 4.1.2 Pengujian Dengan *SMOTE* – Penghapusan *Outliers*

Pada pengujian ini dilakukan teknik *Synthetic Minority Oversampling Technique (SMOTE)* untuk menangani ketidakseimbangan data, setelah proses *SMOTE* dilakukan dilakukan proses penghapusan terhadap *outliers* yang terdeteksi pada proses EDA. Jumlah data yang digunakan untuk pengujian ini setelah dilakukan teknik *SMOTE* dan penghapusan *outliers* adalah 310 data.

Setelah model dilatih, nilai penting setiap fitur diambil dari atribut *feature\_importances\_* yang menunjukkan kontribusi fitur dalam pembagian node di pohon keputusan. Nilai penting dari beberapa model yang dilatih pada subset data berbeda dihitung rata-ratanya untuk mendapatkan estimasi yang lebih stabil. Untuk masing-masing model dilakukan pengujian dengan menggunakan 4 kombinasi fitur yang berbeda berdasarkan fitur penting yang telah ditentukan seperti yang tertera pada gambar (Rohayati & Widani, 2020).



Gambar 4. 2 Fitur Penting Pengujian Dengan *SMOTE* – Penghapusan *Outliers*

#### 4.1.2.1 Model A

Pada model A dengan rasio perbandingan 90:10 digunakan 31 data untuk pengujian. Pada pengujian ini menggunakan *hyperparameter Random Forest* yaitu yaitu  $n\_estimator = 100$ ,  $random\_state = 57$ ,  $bootstrap = true$ , dan  $max\_depth = 50$ . Pada model A dilakukan 4 kali pengujian dengan 4 kombinasi fitur yang berbeda.

##### A. Kombinasi Fitur 1

Kombinasi fitur 1 menggunakan 7 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 1 sebagai berikut.

Tabel 4. 1 Hasil Pengujian Model A Kombinasi Fitur 1

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
233	0	0
58	1	1
84	1	1
184	0	0
62	1	1
...	...	...
123	1	1
255	0	1
197	0	0
12	1	1

<i>ID</i>	<b>Aktual</b>	<b>Hasil</b>
201	0	0

Tabel 4. 2 *Confusion Matrix* Pengujian Model A Kombinasi Fitur 1

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
13	11	4	3

Dari Tabel 4.2 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 13 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 11 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 4 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 3 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{13 + 11}{13 + 11 + 4 + 3} = 0.74$$

Tabel 4. 3 Hasil Evaluasi Pengujian Model A Kombinasi Fitur 1

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.74	0.69	0.79	0.73

## B. Kombinasi Fitur 2

Kombinasi fitur 2 menggunakan 9 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 2 sebagai berikut.

Tabel 4. 4 Hasil Pengujian Model A Kombinasi Fitur 2

<i>ID</i>	<b>Aktual</b>	<b>Hasil</b>
233	0	0
58	1	1
84	1	1

<i>ID</i>	<b>Aktual</b>	<b>Hasil</b>
184	0	0
62	1	1
...	...	...
123	1	1
255	0	0
197	0	0
12	1	1
201	0	0

Tabel 4. 5 *Confusion Matrix* Pengujian Model A Kombinasi Fitur 2

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
13	11	4	3

Dari Tabel 4.5 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 13 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 11 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 4 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 3 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{13 + 11}{13 + 11 + 4 + 3} = 0.77$$

Tabel 4. 6 Hasil Evaluasi Pengujian Model A Kombinasi Fitur 2

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.77	0.73	0.79	0.76

### C. Kombinasi Fitur 3

Kombinasi fitur 3 menggunakan 11 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 3 sebagai berikut.

Tabel 4. 7 Hasil Pengujian Model A Kombinasi Fitur 3

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
233	0	0
58	1	1
84	1	1
184	0	0
62	1	1
...	...	...
123	1	1
255	0	0
197	0	0
12	1	1
201	0	0

Tabel 4. 8 *Confusion Matrix* Pengujian Model A Kombinasi Fitur 3

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
12	11	5	3

Dari Tabel 4.8 model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 12 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 11 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 5 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 3 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{12 + 11}{12 + 11 + 5 + 3} = 0.74$$

Tabel 4. 9 Hasil Evaluasi Pengujian Model A Kombinasi Fitur 3

<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
0.74	0.69	0.79	0.73

#### D. Semua Fitur

Kombinasi fitur 4 menggunakan semua fitur yang ada pada *dataset*. Hasil pengujian yang didapat dengan menggunakan kombinasi semua fitur sebagai berikut.

Tabel 4. 10 Hasil Pengujian Model A Semua Fitur

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
233	0	0
58	1	1
84	1	1
184	0	0
62	1	1
...	...	...
123	1	1
255	0	0
197	0	0
12	1	1
201	0	0

Tabel 4. 11 *Confusion Matrix* Pengujian Model A Semua Fitur

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
13	11	4	3

Dari Tabel 4.11 model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 13 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 11 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 4 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 3 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{13 + 11}{13 + 11 + 4 + 3} = 0.77$$

Tabel 4. 12 Hasil Evaluasi Pengujian Model A Semua Fitur

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.77	0.73	0.79	0.76

#### 4.1.2.2 Model B

Pada model B dengan rasio perbandingan 80:20 digunakan 62 data untuk pengujian. Pada pengujian ini menggunakan *hyperparameter Random Forest* yaitu yaitu  $n\_estimator = 100$ ,  $random\_state = 57$ ,  $bootstrap = true$ , dan  $max\_depth = 50$ . Pada model B dilakukan 4 kali pengujian dengan 4 kombinasi fitur yang berbeda.

##### A. Kombinasi Fitur 1

Kombinasi fitur 1 menggunakan 7 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 1 sebagai berikut.

Tabel 4. 13 Hasil Pengujian Model B Kombinasi Fitur 1

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
233	0	0
58	1	1
84	1	1
184	0	0
62	1	1
...	...	...
7	1	1
186	0	0
87	1	1
305	0	0
138	1	0

Tabel 4. 14 *Confusion Matrix* Pengujian Model B Kombinasi Fitur 1

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
25	25	4	8

Dari Tabel 4.14 model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 25 data dengan nilai 0 (tidak berisiko



serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 25 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 4 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 8 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{25 + 25}{25 + 25 + 4 + 8} = 0.81$$

Tabel 4. 15 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 1

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.81	0.86	0.76	0.81

## B. Kombinasi Fitur 2

Kombinasi fitur 2 menggunakan 9 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 2 sebagai berikut.

Tabel 4. 16 Hasil Pengujian Model B Kombinasi Fitur 2

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
233	0	0
58	1	1
84	1	1
184	0	0
62	1	1
...	...	...
7	1	1
186	0	0
87	1	1
305	0	0
138	1	1

Tabel 4. 17 *Confusion Matrix* Pengujian Model B Kombinasi Fitur 2

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
25	27	4	6

Dari Tabel 4.17 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 25 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 27 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 4 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 6 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{25 + 27}{25 + 27 + 4 + 6} = 0.84$$

Tabel 4. 18 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 2

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.84	0.87	0.82	0.84

### C. Kombinasi Fitur 3

Kombinasi fitur 3 menggunakan 11 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 3 sebagai berikut.

Tabel 4. 19 Hasil Pengujian Model B Kombinasi Fitur 3

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
233	0	0
58	1	1
84	1	1
184	0	0
62	1	1
...	...	...
7	1	1
186	0	0
87	1	1
305	0	0
138	1	0

Tabel 4. 20 *Confusion Matrix* Pengujian Model B Kombinasi Fitur 3

TP	TN	FP	FN
24	27	5	6

Dari Tabel 4.20 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 24 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 27 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 5 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 6 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{24 + 27}{24 + 27 + 5 + 6} = 0.82$$

Tabel 4. 21 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 3

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.82	0.84	0.82	0.83

#### D. Semua Fitur

Kombinasi fitur 4 menggunakan semua fitur yang ada pada *dataset*. Hasil pengujian yang didapat dengan menggunakan kombinasi semua fitur sebagai berikut.

Tabel 4. 22 Hasil Pengujian Model B Semua Fitur

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
233	0	0
58	1	1
84	1	1

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
184	0	0
62	1	1
...	...	...
7	1	1
186	0	0
87	1	1
305	0	0
138	1	0

Tabel 4. 23 *Confusion Matrix* Pengujian Model B Semua Fitur

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
25	27	4	6

Dari Tabel 4.23 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 25 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 27 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 4 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 6 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{25 + 27}{25 + 27 + 4 + 6} = 0.84$$

Tabel 4. 24 Hasil Evaluasi Pengujian Model B Semua Fitur

<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
0.84	0.87	0.82	0.84

#### 4.1.2.3 Model C

Pada model C dengan rasio perbandingan 75:25 digunakan 78 data untuk pengujian. Pada pengujian ini menggunakan *hyperparameter Random Forest*

yaitu yaitu  $n\_estimator = 100$ ,  $random\_state = 57$ ,  $bootstrap = true$ , dan  $max\_depth = 50$ . Pada model C dilakukan 4 kali pengujian dengan 4 kombinasi fitur yang berbeda.

### A. Kombinasi Fitur 1

Kombinasi fitur 1 menggunakan 7 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 1 sebagai berikut.

Tabel 4. 25 Hasil Pengujian Model C Kombinasi Fitur 1

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
233	0	0
58	1	1
84	1	1
184	0	0
62	1	1
...	...	...
301	0	0
115	1	1
91	1	1
275	0	0
151	1	0

Tabel 4. 26 *Confusion Matrix* Pengujian Model C Kombinasi Fitur 1

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
31	32	6	9

Dari Tabel 4.26 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 31 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 32 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 6 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 9 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak

berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{31 + 34}{31 + 34 + 6 + 9} = 0.81$$

Tabel 4. 27 Hasil Evaluasi Pengujian Model C Kombinasi Fitur 1

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.81	0.84	0.79	0.81

## B. Kombinasi Fitur 2

Kombinasi fitur 2 menggunakan 9 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 2 sebagai berikut.

Tabel 4. 28 Hasil Pengujian Model C Kombinasi Fitur 2

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
233	0	0
58	1	1
84	1	1
184	0	0
62	1	1
...	...	...
301	0	0
115	1	1
91	1	1
275	0	0
151	1	0

Tabel 4. 29 *Confusion Matrix* Pengujian Model C Kombinasi Fitur 2

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
30	34	7	7

Dari Tabel 4.29 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 30 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 34 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 7 data dengan nilai 0 (tidak berisiko serangan

jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 7 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{30 + 34}{30 + 34 + 7 + 7} = 0.82$$

Tabel 4. 30 Hasil Evaluasi Pengujian Model C Kombinasi Fitur 2

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.82	0.83	0.83	0.83

### C. Kombinasi Fitur 3

Kombinasi fitur 3 menggunakan 11 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 3 sebagai berikut.

Tabel 4. 31 Hasil Pengujian Model C Kombinasi Fitur 3

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
233	0	0
58	1	1
84	1	1
184	0	0
62	1	1
...	...	...
301	0	0
115	1	1
91	1	1
275	0	0
151	1	0

Tabel 4. 32 *Confusion Matrix* Pengujian Model C Kombinasi Fitur 3

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
31	33	6	8

Dari Tabel 4.32 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 31 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung),

33 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 6 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 8 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{31 + 33}{31 + 33 + 6 + 8} = 0.82$$

Tabel 4. 33 Hasil Evaluasi Pengujian Model C Kombinasi Fitur 3

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.82	0.85	0.80	0.82

#### D. Semua Fitur

Kombinasi fitur 4 menggunakan semua fitur yang ada pada *dataset*. Hasil pengujian yang didapat dengan menggunakan kombinasi semua fitur sebagai berikut.

Tabel 4. 34 Hasil Pengujian Model C Semua Fitur

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
233	0	0
58	1	1
84	1	1
184	0	0
62	1	1
...	...	...
301	0	0
115	1	1
91	1	1
275	0	0
151	1	1

Tabel 4. 35 *Confusion Matrix* Pengujian Model C Semua Fitur

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
28	35	9	6



Dari Tabel 4.35 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 28 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 35 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 9 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 6 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{28 + 35}{28 + 35 + 9 + 6} = 0.81$$

Tabel 4. 36 Hasil Evaluasi Pengujian Model C Semua Fitur

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.81	0.80	0.85	0.82

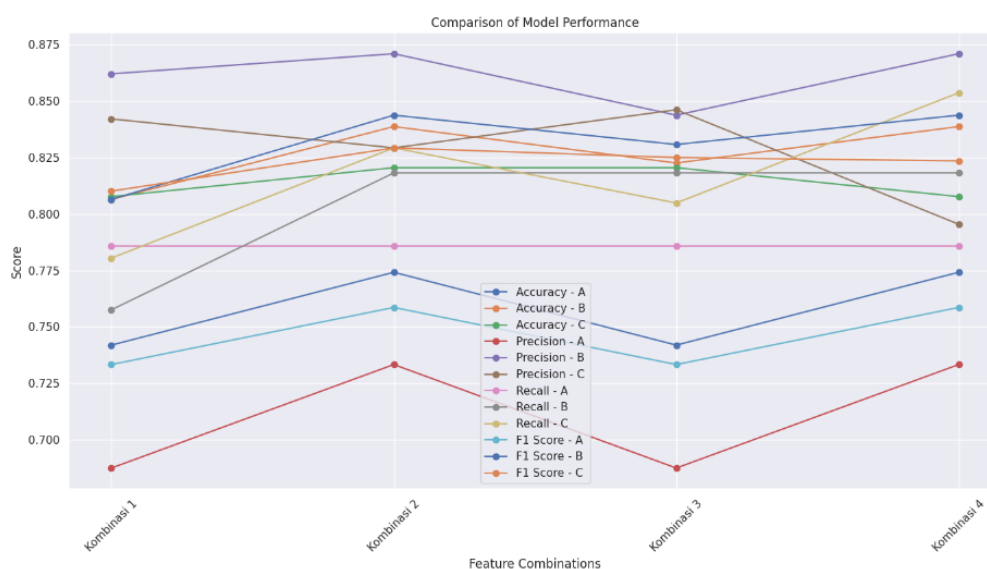
Dari keseluruhan pengujian dengan *SMOTE* – penghapusan *outliers* dapat dilihat hasil matrik evaluasinya pada tabel berikut.

Tabel 4. 37 Hasil Seluruh Pengujian dengan *SMOTE* - Penghapusan Outliers

No	Model	Fitur	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	A	Kombinasi Fitur 1	0.74	0.69	0.79	0.73
2		Kombinasi Fitur 2	0.77	0.73	0.79	0.76
3		Kombinasi Fitur 3	0.74	0.69	0.79	0.73
4		Kombinasi Fitur 4	0.77	0.73	0.79	0.76
5	B	Kombinasi Fitur 1	0.81	0.86	0.76	0.81
6		Kombinasi Fitur 2	0.84	0.87	0.82	0.84
7		Kombinasi Fitur 3	0.82	0.84	0.82	0.83
8		Kombinasi Fitur 4	0.84	0.87	0.82	0.84

No	Model	Fitur	Accuracy	Precision	Recall	F1-Score
9	C	Kombinasi Fitur 1	0.81	0.84	0.79	0.81
10		Kombinasi Fitur 2	0.82	0.83	0.83	0.83
11		Kombinasi Fitur 3	0.82	0.85	0.80	0.82
12		Kombinasi Fitur 4	0.81	0.80	0.85	0.82

Grafik perbandingan hasil evaluasi matrix dari masing-masing model dengan beberapa kelompok kombinasi fitur pada gambar berikut.



Gambar 4. 3 Grafik Evaluasi *Matrix* Pengujian Dengan *SMOTE* – Penghapusan *Outliers*

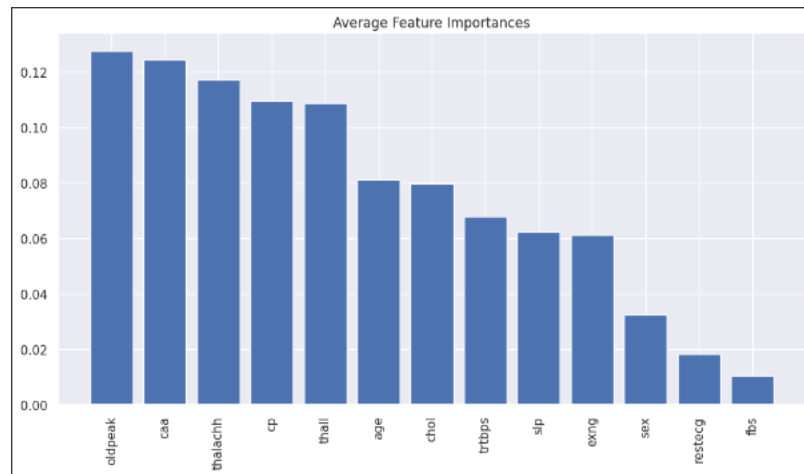
Dari Tabel 4.37 dan Gambar grafik 4.3 perbandingan hasil evaluasi matriks untuk pengujian dengan *SMOTE* – penghapusan *outliers* menunjukkan hasil terbaik pada pengujian model B dengan kombinasi fitur 2 yang menggunakan 9 fitur teratas dari fitur penting, dan kombinasi fitur 4 yang menggunakan seluruh fitur yang ada pada dataset dengan hasil akurasi sebesar 84%.

Pada Gambar 4.3 grafik menunjukkan model B dengan data latih dan uji 80:20 memberikan akurasi terbaik dan paling konsisten di berbagai kombinasi fitur. Hal ini menunjukkan bahwa rasio data yang lebih seimbang antara pelatihan dan pengujian memberikan performa yang lebih baik dan generalisasi yang lebih baik. Kombinasi fitur 2 secara konsisten memberikan akurasi terbaik atau hampir terbaik di semua model. Ini menunjukkan bahwa ada titik optimal dalam jumlah fitur yang digunakan, di mana terlalu sedikit fitur tidak memberikan informasi yang cukup dan terlalu banyak fitur dapat menyebabkan kebisingan atau overfitting.

#### **4.1.3 Pengujian Dengan Penghapusan *Outliers* – *SMOTE***

Pada pengujian ini dilakukan proses penghapusan terhadap *outliers* yang terdeteksi, setelahnya dilakukan teknik *Synthetic Minority Oversampling Technique (SMOTE)* untuk menangani ketidakseimbangan data. Jumlah data yang digunakan untuk pengujian ini setelah dilakukan teknik *SMOTE* dan penghapusan *outliers* adalah 318 data. Pengujian ini dilakukan dengan 3 model pembagian data yaitu model A dengan 90:10, model B dengan 80:20, dan model C dengan 75:25.

Setelah model dilatih, nilai penting setiap fitur diambil dari atribut *feature\_importances\_* yang menunjukkan kontribusi fitur dalam pembagian node di pohon keputusan. Nilai penting dari beberapa model yang dilatih pada subset data berbeda dihitung rata-ratanya untuk mendapatkan estimasi yang lebih stabil. Untuk masing-masing model dilakukan pengujian dengan menggunakan 4 kombinasi fitur yang berbeda berdasarkan fitur penting yang telah ditentukan seperti yang tertera pada gambar (Rohayati & Widani, 2020).



Gambar 4. 4 Fitur Penting Pengujian Dengan Penghapusan *Outliers* – *SMOTE*

#### 4.1.3.1 Model A

Pada model A dengan rasio perbandingan 90:10 digunakan 29 data untuk pengujian. Pada pengujian ini menggunakan *hyperparameter Random Forest* yaitu yaitu  $n\_estimator = 100$ ,  $random\_state = 57$ ,  $bootstrap = true$ , dan  $max\_depth = 50$ . Pada model A dilakukan 4 kali pengujian dengan 4 kombinasi fitur yang berbeda.

##### A. Kombinasi Fitur 1

Kombinasi fitur 1 menggunakan 7 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 1 sebagai berikut.

Tabel 4. 38 Hasil Pengujian Model A Kombinasi Fitur 1

<i>ID</i>	<b>Aktual</b>	<b>Hasil</b>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
171	0	1
295	0	0
143	1	0
263	0	0
129	1	1

Tabel 4. 39 *Confusion Matrix* Pengujian Model A Kombinasi Fitur 1

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
10	13	4	2

Dari Tabel 4.39 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 10 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 13 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 4 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 2 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{10 + 13}{10 + 13 + 4 + 2} = 0.79$$

Tabel 4. 40 Hasil Evaluasi Pengujian Model A Kombinasi Fitur 1

<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
0.79	0.76	0.87	0.81

## B. Kombinasi Fitur 2

Kombinasi fitur 2 menggunakan 9 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 2 sebagai berikut.

Tabel 4. 41 Hasil Pengujian Model A Kombinasi Fitur 2

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
202	0	0
116	1	1
9	1	1
62	0	0
255	1	1
...	...	...

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
171	0	1
295	0	0
143	1	0
263	0	0
129	1	1

Tabel 4. 42 *Confusion Matrix* Pengujian Model A Kombinasi Fitur 2

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
13	13	1	2

Dari Tabel 4.42 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 13 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 13 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 1 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 2 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{13 + 13}{13 + 13 + 1 + 2} = 0.90$$

Tabel 4. 43 Hasil Evaluasi Hasil Pengujian Model A Kombinasi Fitur 2

<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
0.90	0.93	0.87	0.90

### C. Kombinasi Fitur 3

Kombinasi fitur 3 menggunakan 11 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 3 sebagai berikut.

Tabel 4. 44 Hasil Hasil Pengujian Model A Kombinasi Fitur 3

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
202	0	0
116	1	1
9	1	1
62	0	0
255	1	1
...	...	...
171	0	1
295	0	0
143	1	0
263	0	0
129	1	1

Tabel 4. 45 *Confusion Matrix* Hasil Pengujian Model A Kombinasi Fitur 3

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
13	14	1	1

Dari Tabel 4.45 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 13 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 14 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 1 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 1 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{13 + 14}{13 + 14 + 1 + 1} = 0.93$$

Tabel 4. 46 Hasil Evaluasi Hasil Pengujian Model A Kombinasi Fitur 3

<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
0.93	0.93	0.93	0.93

#### D. Semua Fitur

Kombinasi fitur 4 menggunakan semua fitur yang ada pada *dataset*. Hasil pengujian yang didapat dengan menggunakan kombinasi semua fitur sebagai berikut.

Tabel 4. 47 Hasil Pengujian Model A Semua Fitur

<i>ID</i>	<b>Aktual</b>	<b>Hasil</b>
202	0	0
116	1	1
9	1	1
62	0	0
255	1	1
...	...	...
171	0	1
295	0	0
143	1	1
263	0	0
129	1	1

Tabel 4. 48 *Confusion Matrix* Pengujian Model A Semua Fitur

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
13	14	1	1

Dari Tabel 4.48 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 13 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 14 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 1 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 1 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{13 + 14}{13 + 14 + 1 + 1} = 0.93$$



Tabel 4. 49 Hasil Evaluasi Pengujian Model A Semua Fitur

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.93	0.93	0.93	0.93

#### 4.1.3.2 Model B

Pada model B dengan rasio perbandingan 80:20 digunakan 57 data untuk pengujian. Pada pengujian ini menggunakan *hyperparameter Random Forest* yaitu yaitu  $n\_estimator = 100$ ,  $random\_state = 57$ ,  $bootstrap = true$ , dan  $max\_depth = 50$ . Pada model B dilakukan 4 kali pengujian dengan 4 kombinasi fitur yang berbeda.

##### A. Kombinasi Fitur 1

Kombinasi fitur 1 menggunakan 7 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 1 sebagai berikut.

Tabel 4. 50 Hasil Pengujian Model B Kombinasi Fitur 1

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
218	0	0
90	1	1
71	1	1
296	0	1
124	1	1

Tabel 4. 51 *Confusion Matrix* Pengujian Model B Kombinasi Fitur 1

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
19	27	8	3

Dari Tabel 4.51 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 19 data dengan nilai 0 (tidak berisiko

serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 27 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 8 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 3 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{19 + 27}{19 + 27 + 8 + 3} = 0.81$$

Tabel 4. 52 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 1

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.81	0.77	0.90	0.83

## B. Kombinasi Fitur 2

Kombinasi fitur 2 menggunakan 9 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 2 sebagai berikut.

Tabel 4. 53 Hasil Pengujian Model B Kombinasi Fitur 2

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
218	0	0
90	1	1
71	1	1
296	0	1
124	1	1

Tabel 4. 54 *Confusion Matrix* Pengujian Model B Kombinasi Fitur 2

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
20	26	7	4

Dari Tabel 4.54 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 20 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 26 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 7 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 4 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{20 + 26}{20 + 26 + 7 + 4} = 0.81$$

Tabel 4. 55 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 2

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.81	0.79	0.87	0.83

### C. Kombinasi Fitur 3

Kombinasi fitur 3 menggunakan 11 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 3 sebagai berikut.

Tabel 4. 56 Hasil Pengujian Model B Kombinasi Fitur 3

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
202	0	0
116	1	1
9	1	1
62	0	0
255	1	1
...	...	...
218	0	0
90	1	1
71	1	1
296	0	1
124	1	0

Tabel 4. 57 *Confusion Matrix* Pengujian Model B Kombinasi Fitur 3

TP	TN	FP	FN
22	25	5	5

Dari Tabel 4.57 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 22 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 25 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 5 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 5 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{22 + 25}{22 + 25 + 5 + 5} = 0.82$$

Tabel 4. 58 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 3

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.82	0.83	0.83	0.83

#### D. Semua Fitur

Kombinasi fitur 4 menggunakan semua fitur yang ada pada *dataset*. Hasil pengujian yang didapat dengan menggunakan kombinasi semua fitur sebagai berikut.

Tabel 4. 59 Hasil Pengujian Model B Semua Fitur

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
202	0	0
116	1	1
9	1	1

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
62	1	1
255	0	0
...	...	...
218	0	0
90	1	1
71	1	1
296	0	1
124	1	1

Tabel 4. 60 *Confusion Matrix* Model B Semua Fitur

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
22	26	5	4

Dari Tabel 4.60 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 22 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 26 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 5 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 4 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{22 + 26}{22 + 26 + 5 + 4} = 0.84$$

Tabel 4. 61 Hasil Evaluasi Pengujian Model B Semua Fitur

<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
0.84	0.84	0.87	0.85

#### 4.1.3.3 Model C

Pada model C dengan rasio perbandingan 75:25 digunakan 86 data untuk pengujian. Pada pengujian ini menggunakan *hyperparameter Random Forest*

yaitu yaitu  $n\_estimator = 100$ ,  $random\_state = 57$ ,  $bootstrap = true$ , dan  $max\_depth = 50$ . Pada model C dilakukan 4 kali pengujian dengan 4 kombinasi fitur yang berbeda.

### A. Kombinasi Fitur 1

Kombinasi fitur 1 menggunakan 7 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 1 sebagai berikut.

Tabel 4. 62 Hasil Pengujian Model C Kombinasi Fitur 1

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
183	0	0
245	0	1
191	0	0
88	1	1
159	1	1

Tabel 4. 63 *Confusion Matrix* Pengujian Model C Kombinasi Fitur 1

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
25	43	15	3

Dari Tabel 4.63 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 25 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 43 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 15 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 3 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak

berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{25 + 43}{25 + 43 + 15 + 3} = 0.79$$

Tabel 4. 64 Hasil Evaluasi Pengujian Model C Kombinasi Fitur 1

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.79	0.74	0.93	0.83

## B. Kombinasi Fitur 2

Kombinasi fitur 2 menggunakan 9 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 2 sebagai berikut.

Tabel 4. 65 Hasil Pengujian Model C Kombinasi Fitur 2

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
183	0	0
245	0	0
191	0	0
88	1	1
159	1	1

Tabel 4. 66 Confusion Mtrix Pengujian Model C Kombinasi Fitur 2

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
27	44	13	2

Dari Tabel 4.66 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 27 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 44 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 13 data dengan nilai 0 (tidak berisiko serangan

jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 2 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{25 + 43}{25 + 43 + 15 + 3} = 0.82$$

Tabel 4. 67 Hasil Evaluasi Model C Kombinasi Fitur 2

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.83	0.77	0.96	0.85

### C. Kombinasi Fitur 3

Kombinasi fitur 3 menggunakan 11 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 3 sebagai berikut.

Tabel 4. 68 Hasil Pengujian Model C Kombinasi Fitur 3

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
183	0	0
245	0	0
191	0	0
88	1	1
159	1	1

Tabel 4. 69 *Confusion Matrix* Pengujian Model C Kombinasi Fitur 3

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
27	41	13	5

Dari Tabel 4.69 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 27 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung),



41 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 13 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 5 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{27 + 41}{27 + 41 + 13 + 5} = 0.79$$

Tabel 4. 70 Hasil Evaluasi Pengujian Model C Kombinasi Fitur 3

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.79	0.76	0.89	0.82

#### D. Semua Fitur

Kombinasi fitur 4 menggunakan semua fitur yang ada pada *dataset*. Hasil pengujian yang didapat dengan menggunakan kombinasi semua fitur sebagai berikut.

Tabel 4. 71 Hasil Pengujian Model C Semua Fitur

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
183	0	0
245	1	1
191	1	1
88	0	0
159	1	1

Tabel 4. 72 *Confusion Matrix* Pengujian Model C Semua Fitur

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
27	42	13	4

Dari Tabel 4.72 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 27 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 42 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 13 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 4 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{27 + 42}{27 + 42 + 13 + 4} = 0.80$$

Tabel 4. 73 Hasil Evaluasi Pengujian Model C Semua Fitur

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.80	0.76	0.91	0.83

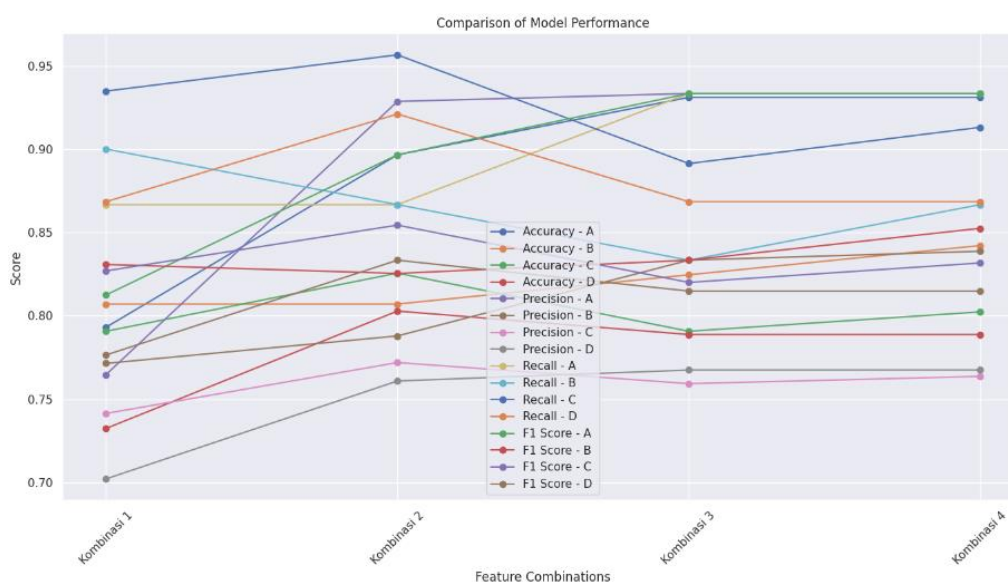
Dari keseluruhan pengujian dengan penghapusan *outliers* – *SMOTE* dapat dilihat hasil matrik evaluasinya pada tabel berikut.

Tabel 4. 74 Hasil Keseluruhan Evaluasi Pengujian Dengan Penghapusan Outliers – *SMOTE*

No	Model	Fitur	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	A	Kombinasi Fitur 1	0.79	0.76	0.87	0.81
2		Kombinasi Fitur 2	0.90	0.93	0.87	0.90
3		Kombinasi Fitur 3	0.93	0.93	0.93	0.93
4		Kombinasi Fitur 4	0.93	0.93	0.93	0.93
5	B	Kombinasi Fitur 1	0.81	0.77	0.90	0.83
6		Kombinasi Fitur 2	0.81	0.79	0.87	0.83
7		Kombinasi Fitur 3	0.82	0.83	0.83	0.83
8		Kombinasi Fitur 4	0.84	0.84	0.87	0.85

No	Model	Fitur	Accuracy	Precision	Recall	F1-Score
9	C	Kombinasi Fitur 1	0.79	0.74	0.93	0.83
10		Kombinasi Fitur 2	0.83	0.77	0.96	0.85
11		Kombinasi Fitur 3	0.79	0.76	0.89	0.82
12		Kombinasi Fitur 4	0.80	0.76	0.91	0.83

Grafik perbandingan hasil evaluasi matrix dari masing-masing model dengan beberapa kelompok kombinasi fitur pada gambar berikut.



Gambar 4. 5 Grafik Evaluasi Matrix Pengujian Dengan Penghapusan *Outliers* – *SMOTE*

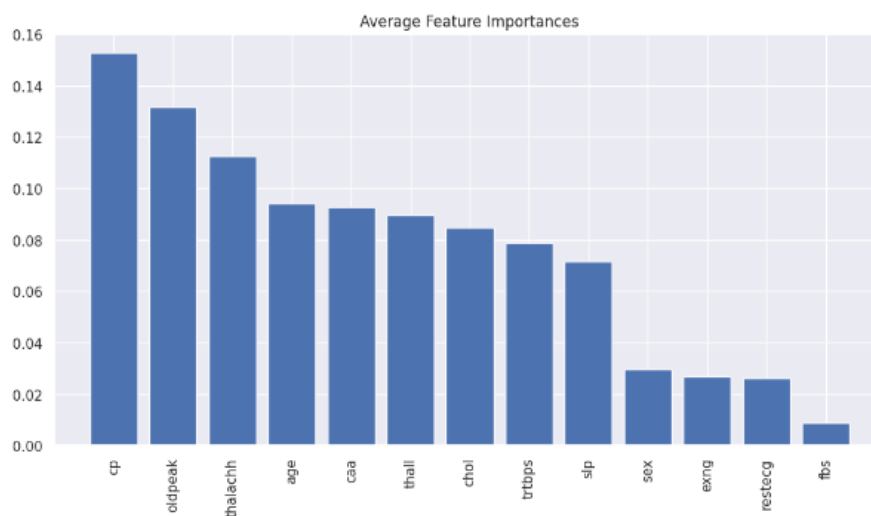
Dari Tabel 4.74 dan Gambar 4.5 grafik perbandingan hasil evaluasi matriks untuk pengujian dengan penghapusan *outliers* – *SMOTE* menunjukkan hasil terbaik pada pengujian model A dengan kombinasi fitur 3 yang menggunakan 11 fitur teratas dari fitur penting, dan kombinasi fitur 4 yang menggunakan seluruh fitur yang ada pada dataset dengan hasil akurasi sebesar 93%.

Pada Gambar 4.5 grafik menunjukkan perbandingan kinerja tiga model yaitu model A, model B, model C dengan kombinasi fitur. Model A menunjukkan peningkatan signifikan pada semua matriks evaluasi dengan penambahan fitur, mencapai puncak pada kombinasi fitur 4. Model B menunjukkan peningkatan yang cenderung stabil dan konsisten terutama pada kombinasi fitur 3 dan 4. Model C mengalami peningkatan signifikan pada kombinasi fitur 2 dan 3 tetapi menurun pada kombinasi fitur 4 yang menunjukkan potensi overfitting. Secara keseluruhan, model A paling responsif terhadap penambahan fitur.

#### **4.1.4 Pengujian Dengan *SMOTE* dan Tanpa Penghapusan *Outliers***

Pada pengujian ini dilakukan proses penghapusan terhadap *outliers* yang terdeteksi. Jumlah data yang digunakan untuk pengujian ini setelah dilakukan teknik *SMOTE* adalah 330 data. Pengujian ini dilakukan dengan 3 model pembagian data yaitu model A dengan 90:10, model B dengan 80:20, dan model C dengan 75:25.

Setelah model dilatih, nilai penting setiap fitur diambil dari atribut *feature\_importances\_* yang menunjukkan kontribusi fitur dalam pembagian node di pohon keputusan. Nilai penting dari beberapa model yang dilatih pada subset data berbeda dihitung rata-ratanya untuk mendapatkan estimasi yang lebih stabil. Untuk masing-masing model dilakukan pengujian dengan menggunakan 4 kombinasi fitur yang berbeda berdasarkan fitur penting yang telah ditentukan seperti yang tertera pada gambar (Rohayati & Widani, 2020).



Gambar 4. 6 Fitur Penting Pengujian Dengan *SMOTE* dan Tanpa Penghapusan *Outliers*

#### 4.1.4.1 Model A

Pada model A dengan rasio perbandingan 90:10 digunakan 29 data untuk pengujian. Pada pengujian ini menggunakan *hyperparameter Random Forest* yaitu yaitu  $n\_estimator = 100$ ,  $random\_state = 57$ ,  $bootstrap = true$ , dan  $max\_depth = 50$ . Pada model A dilakukan 4 kali pengujian dengan 4 kombinasi fitur yang berbeda.

##### A. Kombinasi Fitur 1

Kombinasi fitur 1 menggunakan 7 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 1 sebagai berikut.

Tabel 4. 75 Hasil Pengujian Model A Kombinasi Fitur 1

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
165	0	0
96	1	0
27	1	1
147	1	1
135	1	1
...	...	...
117	1	1
186	0	0
239	0	0
100	1	1

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
118	1	1

Tabel 4. 76 *Confusion Matrix* Pengujian Model A Kombinasi Fitur 1

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
12	16	3	2

Dari Tabel 4.76 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 12 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 16 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 3 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 2 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{12 + 16}{12 + 16 + 3 + 2} = 0.85$$

Tabel 4. 77 Hasil Evaluasi Pengujian Model A Kombinasi 1

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.85	0.84	0.88	0.86

## B. Kombinasi Fitur 2

Kombinasi fitur 2 menggunakan 9 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 2 sebagai berikut.

Tabel 4. 78 Hasil Evaluasi Pengujian Model A Kombinasi Fitur 2

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
165	0	0
96	1	0
27	1	1

<i>ID</i>	<b>Aktual</b>	<b>Hasil</b>
147	1	1
135	1	1
...	...	...
117	1	1
186	0	0
239	0	0
100	1	1
118	1	1

Tabel 4. 79 *Confusion Matrix* Pengujian Model A Kombinasi Fitur 2

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
12	15	3	3

Dari Tabel 4.79 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 12 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 15 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 3 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 3 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{12 + 15}{12 + 15 + 3 + 3} = 0.82$$

Tabel 4. 80 Hasil Evaluasi Pengujian Model A Kombinasi Fitur 2

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.82	0.83	0.83	0.83

### C. Kombinasi Fitur 3

Kombinasi fitur 3 menggunakan 11 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 3 sebagai berikut.

Tabel 4. 81 Hasil Pengujian Model A Kombinasi Fitur 3

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
165	0	0
96	1	0
27	1	1
147	1	1
135	1	1
...	...	...
117	1	1
186	1	1
239	0	0
100	1	1
118	1	1

Tabel 4. 82 *Confusion Matrix* Pengujian Model A Kombinasi Fitur 3

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
12	16	3	2

Dari Tabel 4.82 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 12 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 16 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 3 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 2 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{12 + 16}{12 + 16 + 3 + 2} = 0.85$$

Tabel 4. 83 Hasil Evaluasi Pengujian Model A Kombinasi Fitur 3

<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
0.85	0.84	0.88	0.86



#### D. Semua Fitur

Kombinasi fitur 4 menggunakan semua fitur yang ada pada *dataset*. Hasil pengujian yang didapat dengan menggunakan kombinasi semua fitur sebagai berikut.

Tabel 4. 84 Hasil Pengujian Model A Semua Fitur

<i>ID</i>	<b>Aktual</b>	<b>Hasil</b>
165	0	0
96	1	0
27	1	1
147	1	1
135	1	1
...	...	...
117	1	1
186	0	0
239	0	0
100	1	1
118	1	1

Tabel 4. 85 *Confusion Matrix* Pengujian Model A Semua Fitur

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
14	16	1	2

Dari Tabel 4.85 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 14 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 16 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 1 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 2 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{14 + 16}{14 + 16 + 1 + 2} = 0.91$$

Tabel 4. 86 Hasil Evaluasi Pengujian Model A Semua Fitur

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.91	0.94	0.88	0.91

#### 4.1.4.2 Model B

Pada model B dengan rasio perbandingan 80:20 digunakan 57 data untuk pengujian. Pada pengujian ini menggunakan *hyperparameter Random Forest* yaitu yaitu  $n\_estimator = 100$ ,  $random\_state = 57$ ,  $bootstrap = true$ , dan  $max\_depth = 50$ . Pada model B dilakukan 4 kali pengujian dengan 4 kombinasi fitur yang berbeda.

##### A. Kombinasi Fitur 1

Kombinasi fitur 1 menggunakan 7 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 1 sebagai berikut.

Tabel 4. 87 Hasil Pengujian Model B Kombinasi Fitur 1

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
165	0	0
96	1	0
27	1	1
147	1	1
135	1	1
...	...	...
66	1	1
123	1	1
17	1	0
304	0	0
201	0	0

Tabel 4. 88 *Confusion Matrix* Pengujian Model B Kombinasi Fitur 1

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
25	29	4	8

Dari Tabel 4.88 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 25 data dengan nilai 0 (tidak berisiko

serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 29 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 4 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 8 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{25 + 29}{29 + 25 + 4 + 8} = 0.82$$

Tabel 4. 89 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 1

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.82	0.88	0.78	0.83

## B. Kombinasi Fitur 2

Kombinasi fitur 2 menggunakan 9 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 2 sebagai berikut.

Tabel 4. 90 Hasil Pengujian Model B Kombinasi Fitur 2

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
165	0	0
96	1	0
27	1	1
147	1	1
135	1	1
...	...	...
66	1	1
123	1	1
17	1	1
304	0	0
201	0	0

Tabel 4. 91 *Confusion Matrix* Pengujian Model B Kombinasi Fitur 2

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
26	27	3	10

Dari Tabel 4.91 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 26 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 27 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 3 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 10 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{26 + 27}{26 + 27 + 3 + 10} = 0.80$$

Tabel 4. 92 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 2

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.80	0.90	0.73	0.80

### C. Kombinasi Fitur 3

Kombinasi fitur 3 menggunakan 11 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 3 sebagai berikut.

Tabel 4. 93 Hasil Pengujian Model B Kombinasi Fitur 3

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
165	0	0
96	1	0
27	1	1
147	1	1
135	1	1
...	...	...
66	1	1
123	1	1
17	1	1
304	0	0
201	0	0

Tabel 4. 94 *Confusion Matrix* Pengujian Model B Kombinasi Fitur 3

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
28	27	1	10

Dari Tabel 4.94 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 28 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 27 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 1 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 10 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{28 + 27}{28 + 27 + 1 + 10} = 0.83$$

Tabel 4. 95 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 3

<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
0.83	0.96	0.73	0.83

#### **D. Semua Fitur**

Kombinasi fitur 4 menggunakan semua fitur yang ada pada *dataset*. Hasil pengujian yang didapat dengan menggunakan kombinasi semua fitur sebagai berikut.

Tabel 4. 96 Hasil Pengujian Model B Semua Fitur

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
165	0	0
96	1	0
27	1	1

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
147	1	1
135	1	1
...	...	...
66	1	1
123	1	1
17	1	0
304	0	0
201	0	0

Tabel 4. 97 *Confusion Matrix* Pengujian Model B Semua Fitur

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
26	27	3	10

Dari Tabel 4.97 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 26 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 27 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 3 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 10 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{26 + 27}{26 + 27 + 3 + 10} = 0.80$$

Tabel 4. 98 Hasil Evaluasi Model B Semua Fitur

<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
0.80	0.90	0.73	0.81

#### 4.1.4.3 Model C

Pada model C dengan rasio perbandingan 75:25 digunakan 86 data untuk pengujian. Pada pengujian ini menggunakan *hyperparameter Random Forest*

yaitu yaitu  $n\_estimator = 100$ ,  $random\_state = 57$ ,  $bootstrap = true$ , dan  $max\_depth = 50$ . Pada model C dilakukan 4 kali pengujian dengan 4 kombinasi fitur yang berbeda.

### A. Kombinasi Fitur 1

Kombinasi fitur 1 menggunakan 7 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 1 sebagai berikut.

Tabel 4. 99 Hasil Pengujian Model C Kombinasi Fitur 1

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
165	0	0
96	1	0
27	1	1
147	1	1
135	1	1
...	...	...
32	1	1
137	1	1
180	0	0
230	0	1
249	0	1

Tabel 4. 100 *Confusion Matrix* Model C Kombinasi Fitur 1

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
33	34	7	9

Dari Tabel 4.100 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 33 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 34 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 7 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 9 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak

berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{33 + 34}{33 + 34 + 7 + 9} = 0.81$$

Tabel 4. 101 Hasil Evaluasi Model C Kombinasi Fitur 1

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.81	0.83	0.79	0.81

## B. Kombinasi Fitur 2

Kombinasi fitur 2 menggunakan 9 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 2 sebagai berikut.

Tabel 4. 102 Hasil Pengujian Model C Kombinasi Fitur 2

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
165	0	0
96	1	0
27	1	1
147	1	1
135	1	1
...	...	...
32	1	1
137	1	1
180	0	0
230	0	1
249	0	1

Tabel 4. 103 *Confusion Matrix* Pengujian Model C Kombinasi Fitur 2

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
34	33	6	10

Dari Tabel 4.103 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 34 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 33 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 6 data dengan nilai 0 (tidak berisiko serangan



jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 10 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{34 + 33}{34 + 33 + 6 + 10} = 0.81$$

Tabel 4. 104 Hasil Evaluasi Model C Kombinasi Fitur 2

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.81	0.85	0.77	0.80

### C. Kombinasi Fitur 3

Kombinasi fitur 3 menggunakan 11 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 3 sebagai berikut.

Tabel 4. 105 Hasil Pengujian Model C Kombinasi Fitur 3

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
165	0	0
96	1	0
27	1	1
147	1	1
135	1	1
...	...	...
32	1	1
137	1	1
180	0	0
230	0	1
249	0	0

Tabel 4. 106 *Confusion Matrix* Pengujian Model C Kombinasi Fitur 3

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
37	34	3	9

Dari Tabel 4.106 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 37 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung),

34 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 3 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 9 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{37 + 34}{37 + 34 + 3 + 9} = 0.86$$

Tabel 4. 107 Hasil Evaluasi Matrix Pengujian Model C Kombinasi Fitur 3

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.86	0.92	0.79	0.85

#### D. Semua Fitur

Kombinasi fitur 4 menggunakan semua fitur yang ada pada *dataset*. Hasil pengujian yang didapat dengan menggunakan kombinasi semua fitur sebagai berikut.

Tabel 4. 108 Hasil Pengujian Model C Semua Fitur

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
165	0	0
96	1	0
27	1	1
147	1	1
135	1	1
...	...	...
32	1	1
137	1	1
180	0	0
230	0	1
249	0	0

Tabel 4. 109 *Confusion Matrix* Pengujian Model C Semua Fitur

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
38	33	2	10

Dari Tabel 4.109 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 38 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 33 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 2 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 10 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{38 + 33}{38 + 32 + 2 + 10} = 0.86$$

Tabel 4. 110 Hasil Evaluasi Pengujian Model C Semua Fitur

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.86	0.94	0.77	0.85

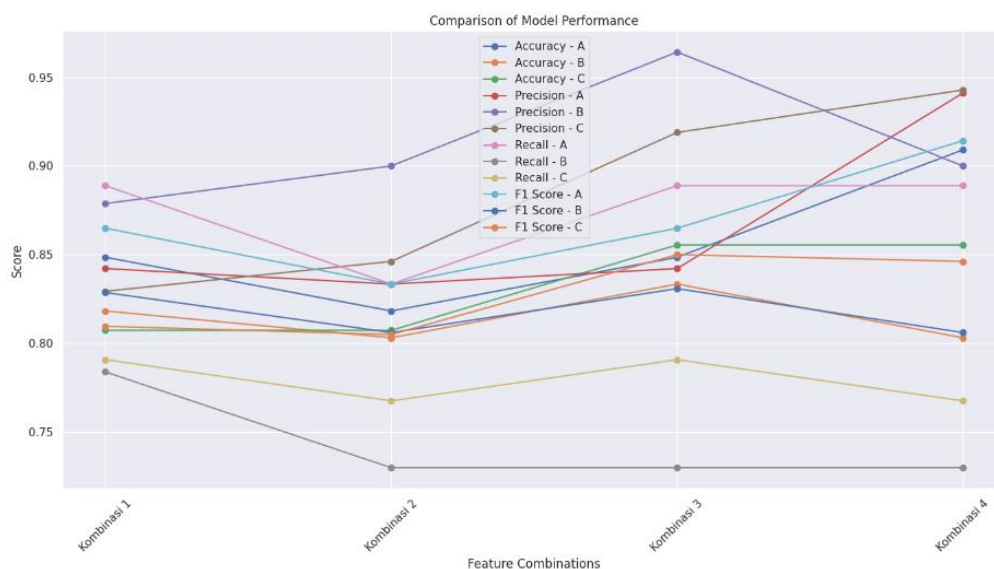
Dari keseluruhan pengujian dengan *SMOTE* dan tanpa penghapusan *outliers* dapat dilihat hasil matrik evaluasinya pada tabel berikut.

Tabel 4. 111 Hasil Evaluasi Keseluruhan Pengujian Dengan *SMOTE* dan Tanpa Penghapusan *Outliers*

No	Model	Fitur	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	A	Kombinasi Fitur 1	0.85	0.84	0.88	0.86
2		Kombinasi Fitur 2	0.82	0.83	0.83	0.83
3		Kombinasi Fitur 3	0.85	0.84	0.88	0.86
4		Kombinasi Fitur 4	0.91	0.94	0.88	0.91
5	B	Kombinasi Fitur 1	0.82	0.88	0.78	0.83
6		Kombinasi Fitur 2	0.80	0.90	0.73	0.80
7		Kombinasi Fitur 3	0.83	0.96	0.73	0.83
8		Kombinasi	0.80	0.90	0.73	0.81

No	Model	Fitur	Accuracy	Precision	Recall	F1-Score
		Fitur 4				
9	C	Kombinasi Fitur 1	0.81	0.83	0.79	0.81
10		Kombinasi Fitur 2	0.81	0.85	0.77	0.80
11		Kombinasi Fitur 3	0.86	0.92	0.79	0.85
12		Kombinasi Fitur 4	0.86	0.94	0.77	0.85

Grafik perbandingan hasil evaluasi matrix dari masing-masing model dengan beberapa kelompok kombinasi fitur pada gambar berikut.



Gambar 4. 7 Grafik Evaluasi Matrix Pengujian Dengan *SMOTE* dan Tanpa Penghapusan *Outliers*

Dari Tabel 4.111 dan Gambar 4.7 grafik perbandingan hasil evaluasi matriks untuk pengujian dengan *SMOTE* dan tanpa penghapusan *outliers* menunjukkan hasil terbaik pada pengujian model A dengan kombinasi fitur 4 yang menggunakan seluruh fitur yang ada pada dataset dengan hasil akurasi sebesar 91%.

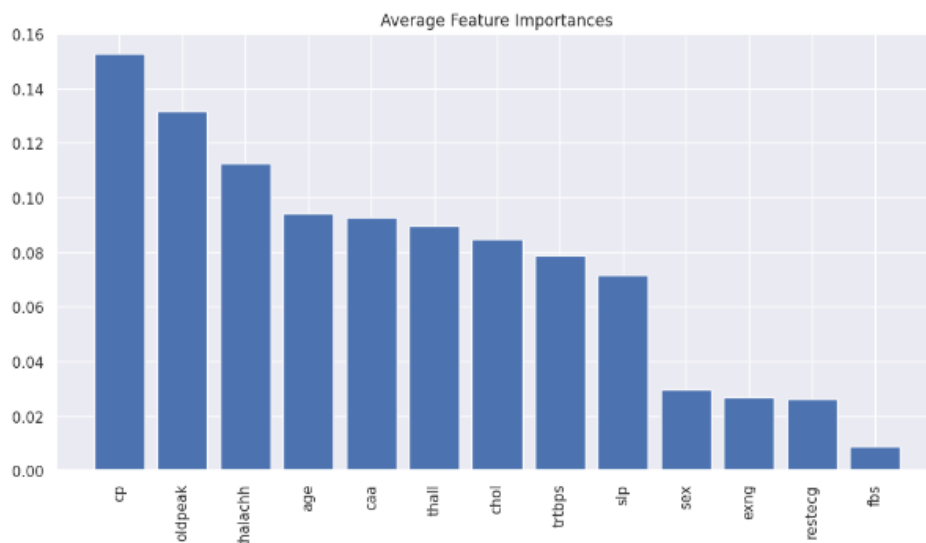
Grafik Gambar 4.7 menunjukkan perbandingan kinerja tiga model yaitu model A, model B, dan model C dengan berbagai kombinasi fitur. Model A

menunjukkan peningkatan performa signifikan dengan penambahan fitur, mencapai hasil terbaik pada kombinasi fitur 4. Model B cenderung stabil tanpa peningkatan signifikan, menunjukkan keterbatasannya dalam memanfaatkan fitur tambahan. Model C menunjukkan peningkatan konsisten dengan penambahan fitur, terutama pada kombinasi fitur 3 dan 4, meskipun tidak setinggi model A. Secara keseluruhan, model A paling responsif terhadap penambahan fitur.

#### **4.1.5 Pengujian Dengan Penghapusan *Outliers* dan Tanpa *SMOTE***

Pada pengujian ini dilakukan proses penghapusan terhadap *outliers* yang terdeteksi. Jumlah data yang digunakan untuk pengujian ini setelah dilakukan teknik *SMOTE* adalah 284 data. Pengujian ini dilakukan dengan 3 model pembagian data yaitu model A dengan 90:10, model B dengan 80:20, dan model C dengan 75:25.

Setelah model dilatih, nilai penting setiap fitur diambil dari atribut *feature\_importances\_* yang menunjukkan kontribusi fitur dalam pembagian node di pohon keputusan. Nilai penting dari beberapa model yang dilatih pada subset data berbeda dihitung rata-ratanya untuk mendapatkan estimasi yang lebih stabil. Untuk masing-masing model dilakukan pengujian dengan menggunakan 4 kombinasi fitur yang berbeda berdasarkan fitur penting yang telah ditentukan seperti yang tertera pada gambar (Rohayati & Widani, 2020).



Gambar 4. 8 Fitur Penting Pengujian Dengan Penghapusan *Outliers* dan Tanpa *SMOTE*

#### 4.1.5.1 Model A

Pada model A dengan rasio perbandingan 90:10 digunakan 29 data untuk pengujian. Pada pengujian ini menggunakan *hyperparameter Random Forest* yaitu yaitu  $n\_estimator = 100$ ,  $random\_state = 57$ ,  $bootstrap = true$ , dan  $max\_depth = 50$ . Pada model A dilakukan 4 kali pengujian dengan 4 kombinasi fitur yang berbeda.

##### A. Kombinasi Fitur 1

Kombinasi fitur 1 menggunakan 7 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 1 sebagai berikut.

Tabel 4. 112 Hasil Pengujian Model A Kombinasi Fitur 1

<i>ID</i>	<b>Aktual</b>	<b>Hasil</b>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
171	0	1
295	0	0
143	1	0
263	1	1

<i>ID</i>	<b>Aktual</b>	<b>Hasil</b>
129	1	1

Tabel 4. 113 *Confusion Matrix* Pengujian Model A Kombinasi Fitur 1

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
12	13	2	2

Dari Tabel 4.113 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 12 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 13 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 2 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 2 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{12 + 13}{12 + 13 + 2 + 2} = 0.85$$

Tabel 4. 114 Hasil Evaluasi Pengujian Model A Kombinasi Fitur 1

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.86	0.87	0.87	0.87

## B. Kombinasi Fitur 2

Kombinasi fitur 2 menggunakan 9 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 2 sebagai berikut.

Tabel 4. 115 Hasil Pengujian Model A Kombinasi Fitur 2

<i>ID</i>	<b>Aktual</b>	<b>Hasil</b>
202	0	0
116	1	1
9	1	1

<i>ID</i>	<b>Aktual</b>	<b>Hasil</b>
62	1	1
255	0	0
...	...	...
171	0	1
295	0	0
143	1	0
263	1	1
129	1	1

Tabel 4. 116 *Confusion Matrix* Model A Kombinasi Fitur 2

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
13	12	1	3

Dari Tabel 4.116 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 13 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 12 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 1 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 3 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{13 + 12}{13 + 12 + 1 + 3} = 0.82$$

Tabel 4. 117 Hasil Evaluasi Model A Kombinasi Fitur 2

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.86	0.92	0.80	0.86

### C. Kombinasi Fitur 3

Kombinasi fitur 3 menggunakan 11 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 3 sebagai berikut.



Tabel 4. 118 Hasil Pengujian Model A Kombinasi Fitur 3

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
171	0	1
295	0	0
143	1	1
263	0	0
129	1	1

Tabel 4. 119 *Confusion Matrix* Pengujian Model A Kombinasi Fitur 3

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
13	13	1	2

Dari Tabel 4.119 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 13 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 13 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 1 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 2 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{13 + 13}{13 + 13 + 1 + 2} = 0.90$$

Tabel 4. 120 Hasil Evaluasi Pengujian Model A Kombinasi Fitur 3

<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
0.90	0.93	0.87	0.90

#### D. Semua Fitur

Kombinasi fitur 4 menggunakan semua fitur yang ada pada *dataset*. Hasil pengujian yang didapat dengan menggunakan kombinasi semua fitur sebagai berikut.

Tabel 4. 121 Hasil Pengujian Model A Semua Fitur

<i>ID</i>	<b>Aktual</b>	<b>Hasil</b>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
171	0	1
295	0	0
143	1	0
263	0	0
129	1	1

Tabel 4. 122 *Confusion Matrix* Pengujian Model A Semua Fitur

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
13	12	1	3

Dari Tabel 4.122 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 13 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 12 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 3 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 2 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{13 + 12}{13 + 12 + 1 + 3} = 0.86$$

Tabel 4. 123 Hasil Evaluasi Model A Semua Fitur

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.86	0.92	0.80	0.86

#### 4.1.5.2 Model B

Pada model B dengan rasio perbandingan 80:20 digunakan 57 data untuk pengujian. Pada pengujian ini menggunakan *hyperparameter Random Forest* yaitu yaitu  $n\_estimator = 100$ ,  $random\_state = 57$ ,  $bootstrap = true$ , dan  $max\_depth = 50$ . Pada model B dilakukan 4 kali pengujian dengan 4 kombinasi fitur yang berbeda.

##### A. Kombinasi Fitur 1

Kombinasi fitur 1 menggunakan 7 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 1 sebagai berikut.

Tabel 4. 124 Hasil Pengujian Model B Kombinasi Fitur 1

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
218	0	0
90	1	1
71	1	1
296	0	1
124	1	1

Tabel 4. 125 *Confusion Matrix* Model B Kombinasi Fitur 1

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
18	27	9	3

Dari Tabel 4.125 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 18 data dengan nilai 0 (tidak berisiko

serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 27 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 9 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 3 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{18 + 27}{18 + 27 + 9 + 3} = 0.79$$

Tabel 4. 126 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 1

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.79	0.75	0.90	0.82

## B. Kombinasi Fitur 2

Kombinasi fitur 2 menggunakan 9 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 2 sebagai berikut.

Tabel 4. 127 Hasil Pengujian Model B Kombinasi Fitur 2

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
218	0	0
90	1	1
71	1	1
296	0	1
124	1	1

Tabel 4. 128 *Confusion Matrix* Model B Kombinasi Fitur 2

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
20	25	7	5

Dari Tabel 4.128 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 20 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 25 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 7 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 5 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{20 + 25}{20 + 25 + 7 + 5} = 0.79$$

Tabel 4. 129 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 2

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.79	0.78	0.83	0.81

### C. Kombinasi Fitur 3

Kombinasi fitur 3 menggunakan 11 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 3 sebagai berikut.

Tabel 4. 130 Hasil Pengujian Model B Kombinasi Fitur 3

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
218	0	0
90	1	1
71	1	1
296	0	1
124	1	1

Tabel 4. 131 *Confusion Matrix* Pengujian Model B Kombinasi Fitur 3

TP	TN	FP	FN
21	27	6	3

Dari Tabel 4.131 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 21 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 27 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 6 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 5 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{21 + 27}{21 + 27 + 6 + 3} = 0.84$$

Tabel 4. 132 Hasil Evaluasi Pengujian Model B Kombinasi Fitur 3

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.84	0.82	0.90	0.86

#### D. Semua Fitur

Kombinasi fitur 4 menggunakan semua fitur yang ada pada *dataset*. Hasil pengujian yang didapat dengan menggunakan kombinasi semua fitur sebagai berikut.

Tabel 4. 133 Hasil Pengujian Model B Semua Fitur

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
202	0	0
116	1	1
9	1	1

<i>ID</i>	<b>Aktual</b>	<b>Hasil</b>
62	1	1
255	0	0
...	...	...
218	0	0
90	1	1
71	1	1
296	0	1
124	1	1

Tabel 4. 134 *Confusion Matrix* Model B Semua Fitur

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
22	26	5	4

Dari Tabel 4.134 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 22 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 26 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 5 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 4 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{22 + 26}{22 + 26 + 5 + 4} = 0.84$$

Tabel 4. 135 Hasil Evaluasi Pengujian Model B Semua Fitur

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.84	0.84	0.87	0.86

#### 4.1.5.3 Model C

Pada model C dengan rasio perbandingan 75:25 digunakan 71 data untuk pengujian. Pada pengujian ini menggunakan *hyperparameter Random Forest*

yaitu yaitu  $n\_estimator = 100$ ,  $random\_state = 57$ ,  $bootstrap = true$ , dan  $max\_depth = 50$ . Pada model C dilakukan 4 kali pengujian dengan 4 kombinasi fitur yang berbeda.

### A. Kombinasi Fitur 1

Kombinasi fitur 1 menggunakan 7 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 1 sebagai berikut.

Tabel 4. 136 Hasil Pengujian Model C Kombinasi Fitur 1

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
83	1	1
131	1	1
166	0	0
146	1	1
115	1	1

Tabel 4. 137 *Confusion Matrix* Pengujian Model C Kombinasi Fitur 1

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
22	34	11	4

Dari Tabel 4.137 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 22 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 34 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 11 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 4 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak



berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{22 + 34}{22 + 34 + 11 + 4} = 0.79$$

Tabel 4. 138 Hasil Evaluasi Pengujian Model C Kombinasi Fitur 1

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.79	0.76	0.89	0.82

## B. Kombinasi Fitur 2

Kombinasi fitur 2 menggunakan 9 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 2 sebagai berikut.

Tabel 4. 139 Hasil Pengujian Model C Kombinasi Fitur 2

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
83	1	0
131	1	1
166	0	0
146	1	1
115	1	1

Tabel 4. 140 *Confusion Matrix* Pengujian Model C Kombinasi Fitur 2

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
22	33	11	5

Dari Tabel 4.140 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 22 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 33 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 11 data dengan nilai 0 (tidak berisiko serangan

jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 5 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{22 + 33}{22 + 33 + 11 + 5} = 0.77$$

Tabel 4. 141 Hasil Evaluasi Pengujian Model C Kombinasi Fitur 2

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.77	0.75	0.87	0.80

### C. Kombinasi Fitur 3

Kombinasi fitur 3 menggunakan 11 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 3 sebagai berikut.

Tabel 4. 142 Hasil Pengujian Model C Kombinasi Fitur 3

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
83	1	0
131	1	1
166	0	0
146	1	1
115	1	1

Tabel 4. 143 *Confusion Matrix* Pengujian Model C Kombinasi Fitur 3

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
23	35	10	3

Dari Tabel 4.143 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 23 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung),

35 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 10 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 3 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{23 + 35}{23 + 35 + 10 + 3} = 0.82$$

Tabel 4. 144 Hasil Evaluasi Pengujian Model C Kombinasi Fitur 3

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.82	0.78	0.92	0.84

#### D. Semua Fitur

Kombinasi fitur 4 menggunakan semua fitur yang ada pada *dataset*. Hasil pengujian yang didapat dengan menggunakan kombinasi semua fitur sebagai berikut.

Tabel 4. 145 Hasil Pengujian Model C Semua Fitur

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
202	0	0
116	1	1
9	1	1
62	1	1
255	0	0
...	...	...
83	1	1
131	1	1
166	0	0
146	1	1
115	1	1

Tabel 4. 146 *Confusion Matrix* Pengujian Model C Semua Fitur

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
24	36	9	2

Dari Tabel 4.146 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 24 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 36 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 9 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 2 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{24 + 36}{24 + 36 + 9 + 2} = 0.85$$

Tabel 4. 147 Hasil Evaluasi Pengujian Model C Semua Fitur

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.85	0.80	0.95	0.87

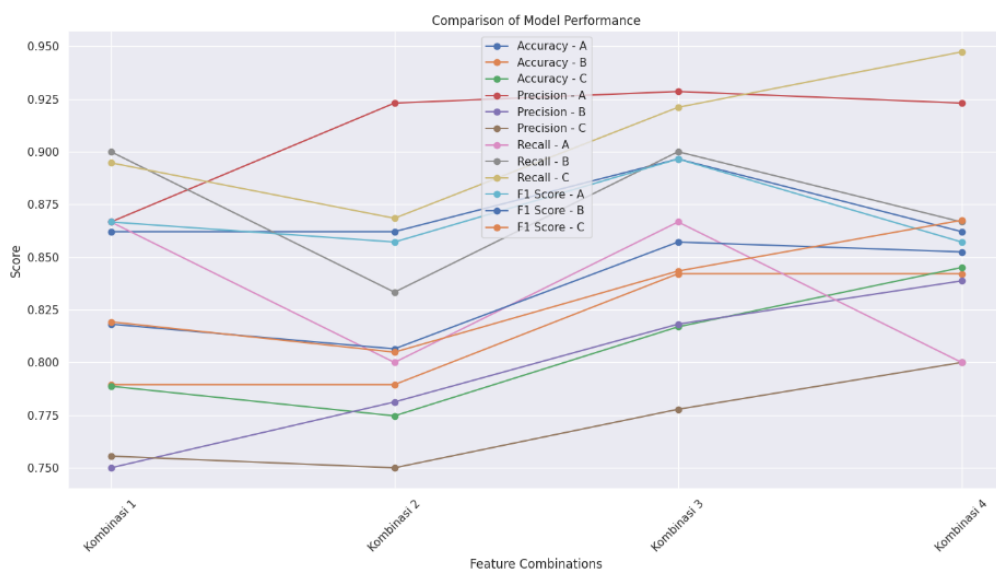
Dari keseluruhan pengujian dengan penghapusan *outliers* dan tanpa *SMOTE* dapat dilihat hasil matrik evaluasinya pada tabel berikut.

Tabel 4. 148 Hasil Evaluasi Seluruh Pengujian Dengan Penghapusan Outliers dan Tanpa *SMOTE*

No	Model	Fitur	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	A	Kombinasi Fitur 1	0.86	0.87	0.87	0.87
2		Kombinasi Fitur 2	0.86	0.92	0.80	0.86
3		Kombinasi Fitur 3	0.90	0.93	0.87	0.90
4		Kombinasi Fitur 4	0.86	0.92	0.80	0.86
5	B	Kombinasi Fitur 1	0.79	0.75	0.90	0.82
6		Kombinasi Fitur 2	0.79	0.78	0.83	0.81
7		Kombinasi Fitur 3	0.84	0.82	0.90	0.86
8		Kombinasi Fitur 4	0.84	0.84	0.87	0.86

No	Model	Fitur	Accuracy	Precision	Recall	F1-Score
9	C	Kombinasi Fitur 1	0.79	0.76	0.89	0.82
10		Kombinasi Fitur 2	0.77	0.75	0.87	0.80
11		Kombinasi Fitur 3	0.82	0.78	0.92	0.84
12		Kombinasi Fitur 4	0.85	0.80	0.95	0.87

Grafik perbandingan hasil evaluasi matrix dari masing-masing model dengan beberapa kelompok kombinasi fitur pada gambar berikut.



Gambar 4. 9 Grafik Evaluasi Pengujian Dengan Penghapusan *Outliers* dan Tanpa *SMOTE*

Dari Tabel 4.148 dan Gambar 4.9 grafik perbandingan hasil evaluasi matriks untuk pengujian dengan penghapusan *outliers* dan tanpa *SMOTE* menunjukkan hasil terbaik pada pengujian model A dengan kombinasi fitur 3 yang menggunakan kombinasi 11 fitur teratas dari fitur penting yang ada pada dataset dengan hasil akurasi sebesar 90%.

Pada Gambar 4.9 grafik menunjukkan model yang paling konsisten dan memberikan hasil terbaik adalah model A, terutama dengan kombinasi fitur 3.

Model B menunjukkan peningkatan akurasi yang stabil dengan penambahan fitur, sementara model C menunjukkan peningkatan signifikan dengan seluruh fitur tetapi kurang konsisten dengan fitur lainnya. Kombinasi fitur 1 memberikan hasil yang cukup baik namun tidak optimal, dan kombinasi fitur 2 tidak efektif serta bisa menurunkan akurasi. Kombinasi fitur 3 memberikan peningkatan akurasi yang signifikan pada semua model, menunjukkan manfaat dari penambahan fitur. Menggunakan seluruh fitur memberikan hasil terbaik untuk beberapa model tetapi tidak selalu meningkatkan akurasi secara signifikan.

#### **4.1.6 Pengujian *Data Real***

Pada pengujian ini digunakan data secara keseluruhan dari *dataset* yang digunakan untuk penelitian ini tanpa dilakukannya *split* data. Pengujian menggunakan *data Real* juga dengan 4 pengujian yaitu pengujian dengan *SMOTE* – penghapusan *outliers*, pengujian dengan penghapusan *outliers* – *SMOTE*, pengujian dengan *SMOTE* dan tanpa penghapusan *outliers*, pengujian dengan penghapusan *outliers* dan tanpa *SMOTE*.

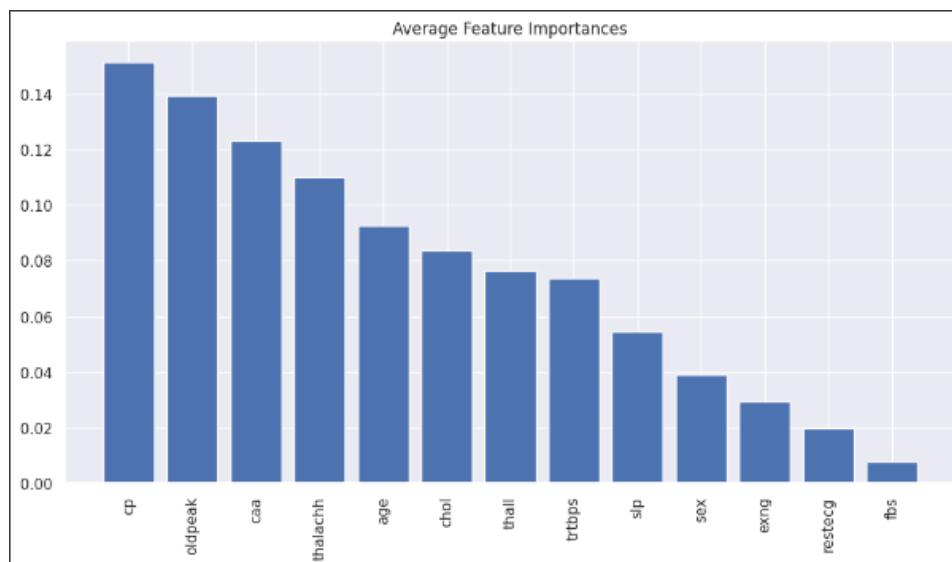
Nilai penting setiap fitur diambil dari atribut *feature\_importances\_* yang menunjukkan kontribusi fitur dalam pembagian node di pohon keputusan. Nilai penting dari beberapa model yang dilatih pada subset data berbeda dihitung rata-ratanya untuk mendapatkan estimasi yang lebih stabil. Untuk masing-masing model dilakukan pengujian dengan menggunakan 4 kombinasi fitur yang berbeda berdasarkan fitur penting yang telah ditentukan (Rohayati & Widani, 2020).

Untuk masing-masing pengujian dilakukan dengan data yang menggunakan 4 kombinasi fitur. Kombinasi fitur 1 dengan menggunakan 7 fitur

teratas dari fitur penting, kombinasi fitur 2 dengan menggunakan 9 fitur teratas dari fitur penting, kombinasi fitur 3 dengan menggunakan 11 fitur teratas dari fitur penting, dan kombinasi fitur 4 dengan menggunakan keseluruhan fitur.

#### 4.1.6.1 Pengujian Dengan *SMOTE* – Penghapusan *Outliers*

Pada pengujian ini digunakan 310 data setelah dilakukan proses *SMOTE* dan penghapusan *outliers*. Pengujian ini menggunakan *hyperparameter Random Forest* yaitu yaitu  $n\_estimator = 100$ ,  $random\_state = 57$ ,  $bootstrap = true$ , dan  $max\_depth = 50$ .



Gambar 4. 10 Fitur Penting Data *Real* Pengujian Dengan *SMOTE* – Penghapusan *Outliers*

#### A. Kombinasi Fitur 1

Kombinasi fitur 1 menggunakan 7 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 1 sebagai berikut.

Tabel 4. 149 Hasil Pengujian *SMOTE* - Penghapusan *Outliers* Data *Real* Kombinasi Fitur 1

<i>ID</i>	Aktual	Hasil
0	1	1
1	1	1
2	1	1
3	1	1

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
4	1	1
...	...	...
324	0	0
326	0	0
327	0	0
328	0	0
329	0	0

Tabel 4. 150 *Confusion Matrix* Pengujian *SMOTE* - Penghapusan Outliers Data *Real* Kombinasi Fitur 1

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
151	159	0	0

Dari Tabel 4.150 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 159 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 159 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 0 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 0 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{151 + 159}{151 + 159 + 0 + 0} = 1.00$$

Tabel 4. 151 Hasil Evaluasi Pengujian *SMOTE* - Penghapusan Outliers Data *Real* Kombinasi Fitur 1

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1.00	1.00	1.00	1.00

## B. Kombinasi Fitur 2



Kombinasi fitur 2 menggunakan 9 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 2 sebagai berikut.

Tabel 4. 152 Hasil Pengujian *SMOTE* - Penghapusan Outliers Data *Real* Kombinasi Fitur 2

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
0	1	1
1	1	1
2	1	1
3	1	1
4	1	1
...	...	...
324	0	0
326	0	0
327	0	0
328	0	0
329	0	0

Tabel 4. 153 *Confusion Matrix* Pengujian *SMOTE* - Penghapusan Outliers Data *Real* Kombinasi Fitur 2

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
151	159	0	0

Dari Tabel 4.153 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 151 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 159 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 0 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 0 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{151 + 159}{151 + 159 + 0 + 0} = 1.00$$

Tabel 4. 154 Hasil Evaluasi Pengujian *SMOTE* - Penghapusan Outliers Data *Real* Kombinasi Fitur 2

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1.00	1.00	1.00	1.00

### C. Kombinasi Fitur 3

Kombinasi fitur 3 menggunakan 11 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 3 sebagai berikut.

Tabel 4. 155 Hasil Pengujian *SMOTE* - Penghapusan Outliers Data *Real* Kombinasi Fitur 3

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
0	1	1
1	1	1
2	1	1
3	1	1
4	1	1
...	...	...
324	0	0
326	0	0
327	0	0
328	0	0
329	0	0

Tabel 4. 156 *Confusion Matrix* Pengujian *SMOTE* - Penghapusan Outliers Data *Real* Kombinasi Fitur 3

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
151	159	0	0

Dari Tabel 4.156 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 151 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 159 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 0 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 0 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak

berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{151 + 159}{151 + 159 + 0 + 0} = 1.00$$

Tabel 4. 157 Hasil Evaluasi Pengujian *SMOTE* - Penghapusan Outliers Data *Real* Kombinasi Fitur 3

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1.00	1.00	1.00	1.00

#### D. Semua Fitur

Kombinasi fitur 4 menggunakan semua fitur yang ada pada *dataset*. Hasil pengujian yang didapat dengan menggunakan kombinasi semua fitur sebagai berikut.

Tabel 4. 158 Hasil Pengujian *SMOTE* - Penghapusan Outliers Data *Real* Semua Fitur

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
0	1	1
1	1	1
2	1	1
3	1	1
4	1	1
...	...	...
324	0	0
326	0	0
327	0	0
328	0	0
329	0	0

Tabel 4. 159 *Confusion Matrix* Pengujian *SMOTE* - Penghapusan Outliers Data *Real* Semua Fitur

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
151	159	0	0

Dari Tabel 4.159 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 151 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung),

159 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 0 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 0 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

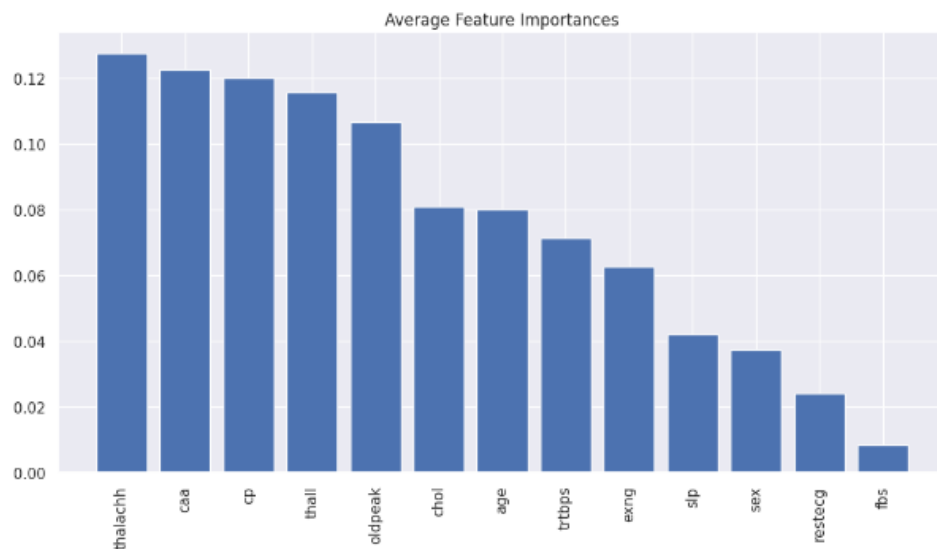
$$Accuracy = \frac{151 + 159}{151 + 159 + 0 + 0} = 1.00$$

Tabel 4. 160 Hasil Evaluasi Pengujian *SMOTE* - Penghapusan Outliers Data *Real* Semua Fitur

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1.00	1.00	1.00	1.00

#### 4.1.6.2 Pengujian Dengan Penghapusan *Outliers* - *SMOTE*

Pada pengujian ini digunakan 284 data setelah dilakukan proses penghapusan *outliers* dan *SMOTE* menggunakan *hyperparameter Random Forest* yaitu yaitu  $n\_estimator = 100$ ,  $random\_state = 57$ ,  $bootstrap = true$ , dan  $max\_depth = 50$ .



Gambar 4. 11 Fitur Penting Data *Real* Pengujian Dengan Penghapusan *Outliers* - *SMOTE*

### A. Kombinasi Fitur 1

Kombinasi fitur 1 menggunakan 7 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 1 sebagai berikut.

Tabel 4. 161 Hasil Pengujian Penghapusan Outliers - *SMOTE Data Real* Kombinasi Fitur 1

<i>ID</i>	<b>Aktual</b>	<b>Hasil</b>
0	1	1
1	1	1
2	1	1
3	1	1
4	1	1
...	...	...
298	0	0
299	0	0
300	0	0
301	0	0
302	0	0

Tabel 4. 162 *Confusion Matrix* Pengujian Penghapusan Outliers - *SMOTE Data Real* Kombinasi Fitur 1

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
125	159	0	0

Dari Tabel 4.162 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 125 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 159 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 0 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 0 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{125 + 159}{125 + 159 + 0 + 0} = 1.00$$

Tabel 4. 163 Hasil Evaluasi Pengujian Penghapusan Outliers - *SMOTE Data Real* Kombinasi Fitur 1

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1.00	1.00	1.00	1.00

## B. Kombinasi Fitur 2

Kombinasi fitur 2 menggunakan 9 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 2 sebagai berikut.

Tabel 4. 164 Hasil Pengujian Penghapusan Outliers - *SMOTE Data Real* Kombinasi Fitur 2

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
0	1	1
1	1	1
2	1	1
3	1	1
4	1	1
...	...	...
298	0	0
299	0	0
300	0	0
301	0	0
302	0	0

Tabel 4. 165 *Confusion Matrix* Pengujian Penghapusan Outliers - *SMOTE Data Real* Kombinasi Fitur 2

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
125	159	0	0

Dari Tabel 4.165 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 125 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 159 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 0 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 0 data

dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{125 + 159}{125 + 159 + 0 + 0} = 1.00$$

Tabel 4. 166 Hasil Evaluasi Pengujian Penghapusan Outliers - *SMOTE Data Real* Kombinasi Fitur 2

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1.00	1.00	1.00	1.00

### C. Kombinasi Fitur 3

Kombinasi fitur 3 menggunakan 11 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 3 sebagai berikut.

Tabel 4. 167 Hasil Pengujian Penghapusan Outliers - *SMOTE Data Real* Kombinasi Fitur 3

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
0	1	1
1	1	1
2	1	1
3	1	1
4	1	1
...	...	...
298	0	0
299	0	0
300	0	0
301	0	0
302	0	0

Tabel 4. 168 *Confusion Matrix* Pengujian Penghapusan Outliers - *SMOTE Data Real* Kombinasi Fitur 3

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
125	159	0	0

Dari Tabel 4.168 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 125 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung),

159 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 0 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 0 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{125 + 159}{125 + 159 + 0 + 0} = 1.00$$

Tabel 4. 169 Hasil Evaluasi Pengujian Penghapusan Outliers - *SMOTE Data Real* Kombinasi Fitur 3

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1.00	1.00	1.00	1.00

#### D. Semua Fitur

Kombinasi fitur 4 menggunakan semua fitur yang ada pada *dataset*. Hasil pengujian yang didapat dengan menggunakan kombinasi semua fitur sebagai berikut.

Tabel 4. 170 Pengujian Penghapusan Outliers - *SMOTE Data Real Semua Fitur*

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
0	1	1
1	1	1
2	1	1
3	1	1
4	1	1
...	...	...
298	0	0
299	0	0
300	0	0
301	0	0
302	0	0

Tabel 4. 171 *Confusion Matrix* Pengujian Penghapusan Outliers - *SMOTE Data Real Semua Fitur*

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
125	159	0	0



Dari Tabel 4.171 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 125 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 159 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 0 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 0 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

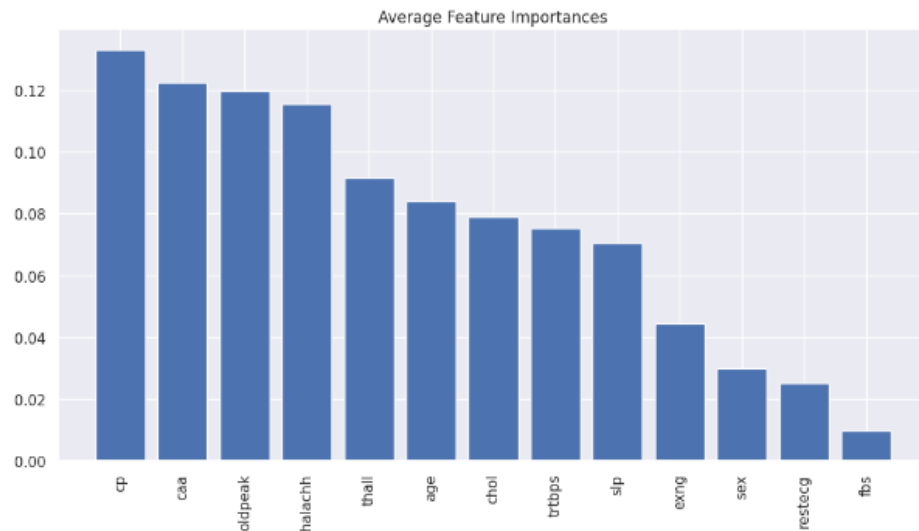
$$Accuracy = \frac{125 + 159}{125 + 159 + 0 + 0} = 1.00$$

Tabel 4. 172 Hasil Evaluasi Pengujian Penghapusan Outliers - *SMOTE* Data *Real* Semua Fitur

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1.00	1.00	1.00	1.00

#### 4.1.6.3 Pengujian Dengan *SMOTE* dan Tanpa Penghapusan *Outliers*

Pada pengujian ini digunakan 330 data setelah dilakukan proses *SMOTE*. Pengujian ini menggunakan *hyperparameter Random Forest* yaitu yaitu  $n\_estimator = 100$ ,  $random\_state = 57$ ,  $bootstrap = true$ , dan  $max\_depth = 50$ .



Gambar 4. 12 Fitur Penting Data *Real* Pengujian Dengan *SMOTE* dan Tanpa Penghapusan *Outliers*

### A. Kombinasi Fitur 1

Kombinasi fitur 1 menggunakan 7 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 1 sebagai berikut.

Tabel 4. 173 Hasil Pengujian Dengan *SMOTE* Kombinasi Fitur 1

<i>ID</i>	<b>Aktual</b>	<b>Hasil</b>
0	1	1
1	1	1
2	1	1
3	1	1
4	1	1
...	...	...
325	0	0
326	0	0
327	0	0
328	0	0
329	0	0

Tabel 4. 174 *Confusion Matrix* Pengujian Dengan *SMOTE* Data *Real* Kombinasi Fitur 1

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
165	165	0	0

Dari Tabel 4.174 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 165 data dengan nilai 0 (tidak berisiko

serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 165 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 0 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 0 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{165 + 165}{165 + 165 + 0 + 0} = 1.00$$

Tabel 4. 175 Hasil Evaluasi Pengujian Dengan *SMOTE* Data *Real* Kombinasi Fitur 1

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1.00	1.00	1.00	1.00

## B. Kombinasi Fitur 2

Kombinasi fitur 2 menggunakan 9 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 2 sebagai berikut.

Tabel 4. 176 Hasil Pengujian Dengan *SMOTE* Data *Real* Kombinasi Fitur 2

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
0	1	1
1	1	1
2	1	1
3	1	1
4	1	1
...	...	...
325	0	0
326	0	0
327	0	0
328	0	0
329	0	0

Tabel 4. 177 Hasil Evaluasi Pengujian Dengan Data *Real SMOTE* Kombinasi Fitur 2

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
165	165	0	0

Dari Tabel 4.177 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 165 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 165 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 0 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 0 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{165 + 165}{165 + 165 + 0 + 0} = 1.00$$

Tabel 4. 178 Hasil Evaluasi Pengujian Dengan *SMOTE Data Real* Kombinasi Fitur 2

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1.00	1.00	1.00	1.00

### C. Kombinasi Fitur 3

Kombinasi fitur 3 menggunakan 11 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 3 sebagai berikut.

Tabel 4. 179 Hasil Pengujian Dengan *SMOTE Data Real* Kombinasi Fitur 3

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
0	1	1
1	1	1
2	1	1
3	1	1
4	1	1
...	...	...
325	0	0
326	0	0
327	0	0
328	0	0
329	0	0

Tabel 4. 180 *Confusion Matrix* Pengujian Dengan *SMOTE Data Real* Kombinasi Fitur 3

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
165	165	0	0

Dari Tabel 4.180 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 165 data *true positive*, 165 data *true negative*, 0 data *false positive*, dan 0 data *false negative*. Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{165 + 165}{165 + 165 + 0 + 0} = 1.00$$

Tabel 4. 181 Hasil Evaluasi Pengujian Dengan *SMOTE Data Real* Kombinasi Fitur 3

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1.00	1.00	1.00	1.00

#### D. Semua Fitur

Kombinasi fitur 4 menggunakan semua fitur yang ada pada *dataset*. Hasil pengujian yang didapat dengan menggunakan kombinasi semua fitur sebagai berikut.

Tabel 4. 182 Hasil Pengujian Dengan *SMOTE Data Real* Semua Fitur

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
0	1	1
1	1	1
2	1	1
3	1	1
4	1	1
...	...	...
325	0	0
326	0	0
327	0	0
328	0	0
329	0	0

Tabel 4. 183 *Confusion Matrix* Pengujian Dengan *SMOTE* Data *Real* Semua Fitur

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
165	165	0	0

Dari Tabel 4.183 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 165 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 165 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 0 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 0 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

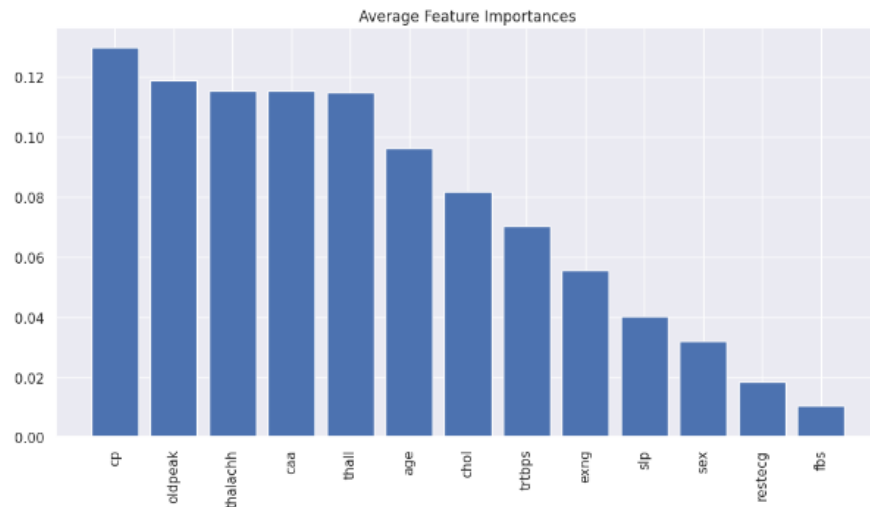
$$Accuracy = \frac{165 + 165}{165 + 165 + 0 + 0} = 1.00$$

Tabel 4. 184 Hasil Evaluasi Pengujian Dengan *SMOTE* Data *Real* Semua Fitur

<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
1.00	1.00	1.00	1.00

#### 4.1.6.4 Pengujian Dengan Penghapusan *Outliers* dan Tanpa *SMOTE*

Pada pengujian ini digunakan 284 data setelah dilakukan proses penghapusan *outliers*. Pengujian ini menggunakan *hyperparameter Random Forest* yaitu yaitu  $n\_estimator = 100$ ,  $random\_state = 57$ ,  $bootstrap = true$ , dan  $max\_depth = 50$ .



Gambar 4. 13 Fitur Penting Data *Real* Pengujian Dengan Penghapusan *Outliers* dan Tanpa *SMOTE*

### A. Kombinasi Fitur 1

Kombinasi fitur 1 menggunakan 7 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 1 sebagai berikut.

Tabel 4. 185 Hasil Pengujian Dengan Penghapusan *Outliers* Data *Real* Kombinasi Fitur 1

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
0	1	1
1	1	1
2	1	1
3	1	1
4	1	1
...	...	...
298	0	0
299	0	0
300	0	0
301	0	0
302	0	0

Tabel 4. 186 *Confusion Matrix* Pengujian Dengan Penghapusan *Outliers* Data *Real* Kombinasi Fitur 1

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
125	159	0	0

Dari Tabel 4.186 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 125 data dengan nilai 0 (tidak berisiko

serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 159 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 0 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 0 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{125 + 159}{125 + 159 + 0 + 0} = 1.00$$

Tabel 4. 187 Hasil Evaluasi Pengujian Dengan Penghapusan Outliers Data *Real* Kombinasi Fitur 1

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1.00	1.00	1.00	1.00

## B. Kombinasi Fitur 2

Kombinasi fitur 2 menggunakan 9 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 2 sebagai berikut.

Tabel 4. 188 Hasil Pengujian Dengan Penghapusan Outliers Data *Real* Kombinasi Fitur 2

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
0	1	1
1	1	1
2	1	1
3	1	1
4	1	1
...	...	...
298	0	0
299	0	0
300	0	0
301	0	0
302	0	0

Tabel 4. 189 *Confusion Matrix* Pengujian Dengan Penghapusan Outliers Data *Real* Kombinasi Fitur 2

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
125	159	0	0



Dari Tabel 4.189 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 125 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 159 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 0 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 0 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{125 + 159}{125 + 159 + 0 + 0} = 1.00$$

Tabel 4. 190 Hasil Evaluasi Pengujian Dengan Penghapusan Outliers Data *Real* Kombinasi Fitur 2

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1.00	1.00	1.00	1.00

### C. Kombinasi Fitur 3

Kombinasi fitur 3 menggunakan 11 fitur teratas dari fitur penting. Hasil pengujian yang didapat dengan menggunakan kombinasi fitur 3 sebagai berikut.

Tabel 4. 191 Hasil Pengujian Dengan Penghapusan Outliers Data *Real* Kombinasi Fitur 3

<i>ID</i>	<i>Aktual</i>	<i>Hasil</i>
0	1	1
1	1	1
2	1	1
3	1	1
4	1	1
...	...	...
298	0	0
299	0	0
300	0	0
301	0	0
302	0	0

Tabel 4. 192 *Confusion Matrix* Pengujian Dengan Penghapusan Outliers Data *Real* Kombinasi Fitur 3

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
125	159	0	0

Dari Tabel 4.192 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 125 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 159 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 0 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 0 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{125 + 159}{125 + 159 + 0 + 0} = 1.00$$

Tabel 4. 193 Hasil Evaluasi Pengujian Dengan Penghapusan Outliers Data *Real* Kombinasi Fitur 3

<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
1.00	1.00	1.00	1.00

#### D. Semua Fitur

Kombinasi fitur 4 menggunakan semua fitur yang ada pada *dataset*. Hasil pengujian yang didapat dengan menggunakan kombinasi semua fitur sebagai berikut.

Tabel 4. 194 Hasil Pengujian Dengan Penghapusan Outliers Data *Real* Semua Fitur

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
0	1	1
1	1	1

<b>ID</b>	<b>Aktual</b>	<b>Hasil</b>
2	1	1
3	1	1
4	1	1
...	...	...
298	0	0
299	0	0
300	0	0
301	0	0
302	0	0

Tabel 4. 195 *Confusion Matrix* Pengujian Dengan Penghapusan Outliers Data *Real* Semua Fitur

<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
125	159	0	0

Dari Tabel 4.195 menunjukkan model *Random Forest Classifier* dalam mengklasifikasikan data menghasilkan 125 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 0 (tidak berisiko serangan jantung), 159 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), 0 data dengan nilai 0 (tidak berisiko serangan jantung) dan hasil aktualnya adalah 1 (berisiko serangan jantung), dan 0 data dengan nilai 1 (berisiko serangan jantung) dan hasil aktualnya data 0 (tidak berisiko serangan jantung). Hasil matrik evaluasi untuk pengujian ini adalah sebagai berikut.

$$Accuracy = \frac{125 + 159}{125 + 159 + 0 + 0} = 1.00$$

Tabel 4. 196 Hasil Evaluasi Pengujian Dengan Penghapusan Outliers Data *Real* Semua Fitur

<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
1.00	1.00	1.00	1.00

Dari keseluruhan pengujian dengan *data Real* dapat dilihat hasil matrik evaluasinya pada tabel berikut.

Tabel 4.197 Hasil Evaluasi Keseluruhan Pengujian Data *Real*

No	Pengujian	Fitur	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	<i>SMOTE</i> – Penghapusan <i>Outliers</i>	Kombinasi Fitur 1	1.00	1.00	1.00	1.00
2		Kombinasi Fitur 2	1.00	1.00	1.00	1.00
3		Kombinasi Fitur 3	1.00	1.00	1.00	1.00
4		Kombinasi Fitur 4	1.00	1.00	1.00	1.00
5	Penghapusan <i>Outliers</i> - <i>SMOTE</i>	Kombinasi Fitur 1	1.00	1.00	1.00	1.00
6		Kombinasi Fitur 2	1.00	1.00	1.00	1.00
7		Kombinasi Fitur 3	1.00	1.00	1.00	1.00
8		Kombinasi Fitur 4	1.00	1.00	1.00	1.00
9	<i>SMOTE</i>	Kombinasi Fitur 1	1.00	1.00	1.00	1.00
10		Kombinasi Fitur 2	1.00	1.00	1.00	1.00
11		Kombinasi Fitur 3	1.00	1.00	1.00	1.00
12		Kombinasi Fitur 4	1.00	1.00	1.00	1.00
13	Penghapusan <i>Outliers</i>	Kombinasi Fitur 1	1.00	1.00	1.00	1.00
14		Kombinasi Fitur 2	1.00	1.00	1.00	1.00
15		Kombinasi Fitur 3	1.00	1.00	1.00	1.00
16		Kombinasi Fitur 4	1.00	1.00	1.00	1.00

Dari Tabel 4.197 perbandingan hasil evaluasi matriks untuk pengujian dengan *data Real*, algoritma *Random Forest* menunjukkan hasil yang sangat baik untuk setiap pengujiannya. Hasil yang didapatkan dari keseluruhan pengujian sesuai dengan data aktualnya, diketahui dari nilai akurasi, *precision*, *recall*, dan *f1 – score* yang menunjukkan nilai 100%.

## 4.2 Pembahasan

Pada bagian ini menjabarkan dari hasil uji coba untuk mengetahui skenario yang paling optimal dalam melatih model *Random Forest*. *Dataset* yang digunakan pada penelitian ini adalah *Heart Attack Analysis & Prediction Dataset* yang merupakan *open access* data web dari Kaggle dengan alamat [kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset/data](https://kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset/data). Total data yang didapat berjumlah 303 data dengan 14 atribut. Pada penelitian ini digunakan 13 fitur karena 1 atributnya yaitu *Num* yang merupakan variabel output.

### 4.2.1 Pengujian Dengan *SMOTE* – Penghapusan *Outliers*

Pada variabel target terdapat *imbalanced* data antara kelas 0 (tidak berisiko serangan jantung) memiliki 138 data, sedangkan kelas 1 (berisiko serangan jantung) memiliki 165 data. Meskipun memiliki perbedaan yang tidak begitu signifikan, namun peneliti tetap melakukan teknik *SMOTE* untuk menangani ketidakseimbangan data tersebut.

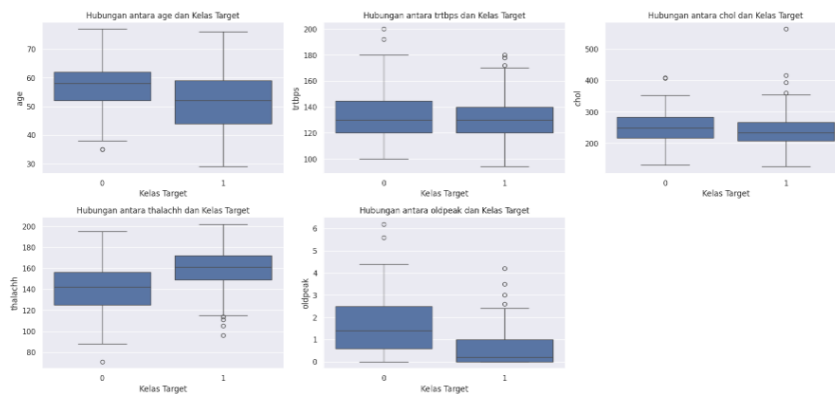


Gambar 4. 14 Grafik Perbandingan Sebelum dan Setelah *SMOTE*

Gambar 4.14 menunjukkan bahwa data target pada *dataset Heart Attack Analysis & Prediction Dataset* sudah seimbang atau *balanced* setelah dilakukan

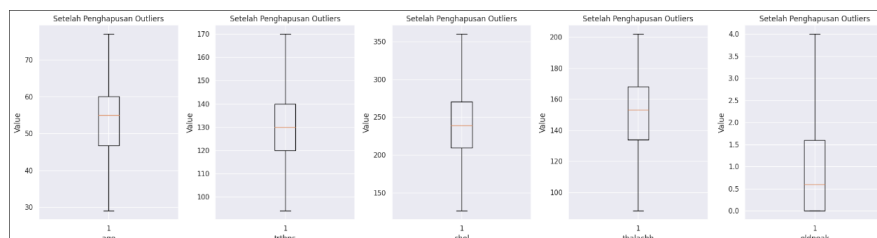
*preprocessing* dengan *Synthetic Minority Oversampling Technique (SMOTE)*. Setelah dilakukan proses *SMOTE* selanjutnya dilakukan *Exploratory Data Analysis (EDA)* untuk memahami hubungan dari masing-masing variabel dengan kelas target, dan mencari *outliers* pada variabel numerical untuk mendapatkan hasil yang lebih optimal.

Proses EDA dilakukan untuk variabel numerical dengan menggunakan *boxplot* untuk menunjukkan hubungan setiap variabel numerical dengan kelas target, serta untuk melihat adanya *outliers* pada setiap variabel numerical.



Gambar 4. 15 *Outliers* yang Terdeteksi

Pada Gambar 4.15 terdapat beberapa *outliers* yang terdeteksi pada proses EDA. Selanjutnya dilakukan penanganan *outliers* yang terdeteksi pada proses EDA tersebut dengan melakukan penghapusan *outliers* yang terdeteksi.



Gambar 4. 16 Setelah Proses Penghapusan *Outliers*

Setelah dilakukan penanganan untuk *outlier* yang terdeteksi, jumlah data yang akan digunakan untuk pengujian adalah 310 data. Selanjutnya adalah membagi data latih dan data uji menjadi 3 model. Model A dengan perbandingan data latih (*train*) sebesar 90%, dan data uji (*test*) sebesar 10%. Model B dengan perbandingan data latih (*train*) sebesar 80%, dan data uji (*test*) sebesar 20%. Model C dengan perbandingan data latih (*train*) sebesar 75%, dan data uji (*test*) sebesar 25% (Sumwiza et al., 2023) dengan menggunakan 4 data kombinasi fitur yang berbeda.

Setelah pembagian data latih dan data uji menjadi 3 model dengan kombinasi 4 fitur yang berbeda, selanjutnya dilakukan *tuning hyperparameter* untuk setiap model yang ada dengan mengatur *n\_estimator* serta *max\_depth* untuk pengujian dengan *SMOTE* – penghapusan *outliers*. Hasil terbaik dari pengujian dengan *SMOTE* – penghapusan *outliers* tertera pada tabel berikut.

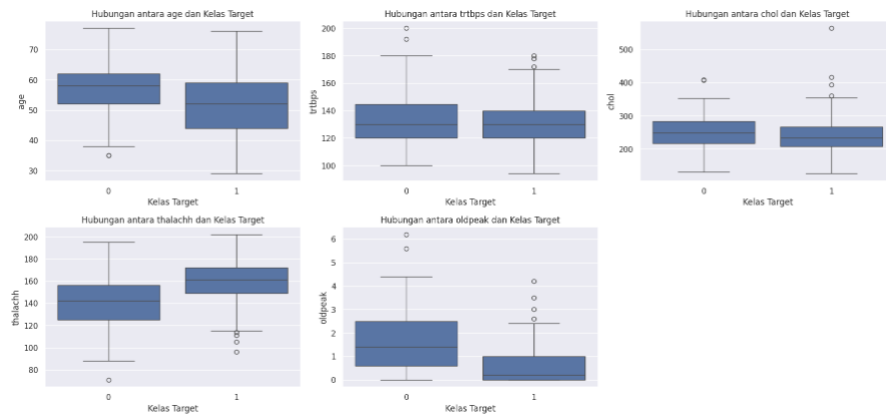
Tabel 4. 198 Hasil Terbaik Dari Pengujian Dengan *SMOTE* – Penghapusan Outliers

No	Model	Fitur	Accuracy	Precision	Recall	F1-Score
1	Model B	Kombinasi Fitur 2	0.84	0.87	0.82	0.84
2	Model B	Kombinasi Fitur 4	0.84	0.87	0.82	0.84

#### 4.2.2 Pengujian Dengan Penghapusan *Outliers* – *SMOTE*

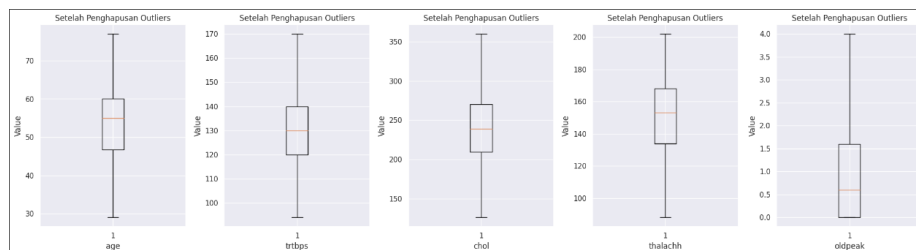
*Exploratory Data Analysis* (EDA) digunakan untuk memahami hubungan dari masing-masing variabel dengan kelas target, dan mencari *outliers* pada variabel numerical untuk mendapatkan hasil yang lebih optimal. Proses EDA dilakukan untuk variabel numerikal dengan menggunakan *boxplot* untuk

menunjukkan hubungan setiap variabel numerikal dengan kelas target, serta untuk melihat adanya *outliers* pada setiap variabel numerikal.



Gambar 4. 17 *Outliers* yang Terdeteksi

Pada Gambar 4.17 terdapat beberapa *outliers* yang terdeteksi pada proses EDA. Selanjutnya dilakukan penanganan *outliers* yang terdeteksi pada proses EDA tersebut dengan melakukan penghapusan *outliers* yang terdeteksi.



Gambar 4. 18 Setelah Proses Penanganan *Outliers*

Setelah dilakukan penanganan untuk *outlier* yang terdeteksi, jumlah data yang akan digunakan untuk pengujian adalah 284 data. Selanjutnya dilakukan proses *Synthetic Minority Oversampling Technique (SMOTE)* untuk menangani ketidakseimbangan data. Pada variabel target terdapat *imbalanced* data antara kelas 0 (tidak berisiko serangan jantung) memiliki 125 data, sedangkan kelas 1 (berisiko serangan jantung) memiliki 159 data. Meskipun memiliki perbedaan



yang tidak begitu signifikan, namun peneliti tetap melakukan teknik *SMOTE* untuk menangani ketidakseimbangan data tersebut.



Gambar 4. 19 Grafik Data Sebelum *SMOTE*



Gambar 4. 20 Grafik Data Setelah *SMOTE*

Gambar 4.20 menunjukkan bahwa data target pada *dataset Heart Attack Analysis & Prediction Dataset* sudah seimbang atau *balanced* setelah dilakukan *preprocessing* menggunakan teknik *Synthetic Minority Oversampling Technique (SMOTE)* dengan masing-masing kelas target yaitu kelas 0 (tidak berisiko serangan jantung) dan kelas 1 (berisiko serangan jantung) berjumlah 159 data.

Selanjutnya adalah membagi data latih dan data uji menjadi 3 model. Model A dengan perbandingan data latih (*train*) sebesar 90%, dan data uji (*test*) sebesar 10%. Model B dengan perbandingan data latih (*train*) sebesar 80%, dan data uji (*test*) sebesar 20%. Model C dengan perbandingan data latih (*train*) sebesar 75%, dan data uji (*test*) sebesar 25% (Sumwiza et al., 2023) dengan menggunakan 4 data kombinasi fitur yang berbeda.

Setelah pembagian data latih dan data uji menjadi 3 model dengan kombinasi 4 fitur yang berbeda, selanjutnya dilakukan *tuning hyperparameter* untuk setiap model yang ada dengan mengatur *n\_estimator* serta *max\_depth* untuk pengujian dengan penghapusan *outliers* – *SMOTE*. Hasil terbaik dari pengujian dengan penghapusan *outliers* – *SMOTE* tertera pada tabel berikut.

Tabel 4. 199 Hasil Terbaik Dari Pengujian Dengan Penghapusan Outliers – *SMOTE*

No	Model	Fitur	Accuracy	Precision	Recall	F1-Score
1	Model A	Kombinasi Fitur 3	0.93	0.93	0.93	0.93
2	Model A	Kombinasi Fitur 4	0.93	0.93	0.93	0.93

#### 4.2.1 Pengujian Dengan *SMOTE* dan Tanpa Penghapusan *Outliers*

Pada variabel target terdapat *imbalanced* data antara kelas 0 (tidak berisiko serangan jantung) memiliki 138 data, sedangkan kelas 1 (berisiko serangan jantung) memiliki 165 data. Meskipun memiliki perbedaan yang tidak begitu signifikan, namun peneliti tetap melakukan teknik *SMOTE* untuk menangani ketidakseimbangan data tersebut.



Gambar 4. 21 Grafik Data Sebelum dan Setelah *SMOTE*

Gambar 4.21 menunjukkan bahwa data target pada *dataset Heart Attack Analysis & Prediction Dataset* sudah seimbang atau *balanced* setelah dilakukan *preprocessing* menggunakan *Synthetic Minority Oversampling Technique (SMOTE)* dengan masing-masing kelas target yaitu kelas 0 (tidak berisiko serangan jantung) dan kelas 1 (berisiko serangan jantung) berjumlah 165 data.

Setelah dilakukan proses *SMOTE*, jumlah data yang akan digunakan untuk pengujian adalah 330 data. Selanjutnya adalah membagi data latih dan data uji menjadi 3 model. Model A dengan perbandingan data latih (*train*) sebesar 90%, dan data uji (*test*) sebesar 10%. Model B dengan perbandingan data latih (*train*) sebesar 80%, dan data uji (*test*) sebesar 20%. Model C dengan perbandingan data latih (*train*) sebesar 75%, dan data uji (*test*) sebesar 25% (Sumwiza et al., 2023) dengan menggunakan 4 data kombinasi fitur yang berbeda.

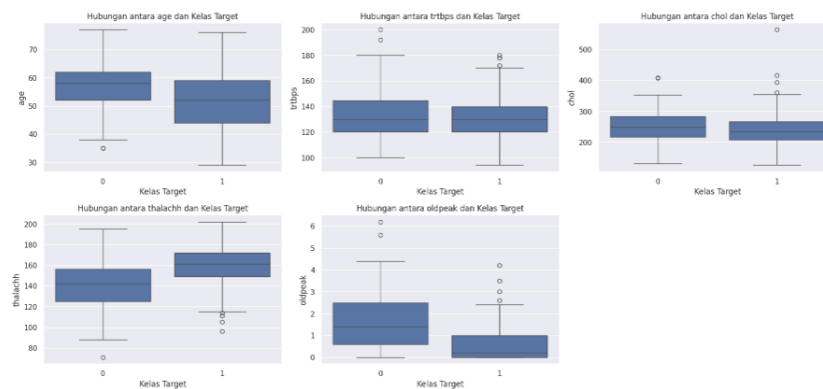
Setelah pembagian data latih dan data uji menjadi 3 model dengan kombinasi 4 fitur yang berbeda, selanjutnya dilakukan *tuning hyperparameter* untuk setiap model yang ada dengan mengatur *n\_estimator* serta *max\_depth* untuk pengujian dengan *SMOTE* – penghapusan *outliers*. Hasil terbaik dari pengujian dengan *SMOTE* – penghapusan *outliers* tertera pada tabel berikut.

Tabel 4. 200 Hasil Terbaik Dari Pengujian Dengan *SMOTE* Tanpa Penghapusan *Outliers*

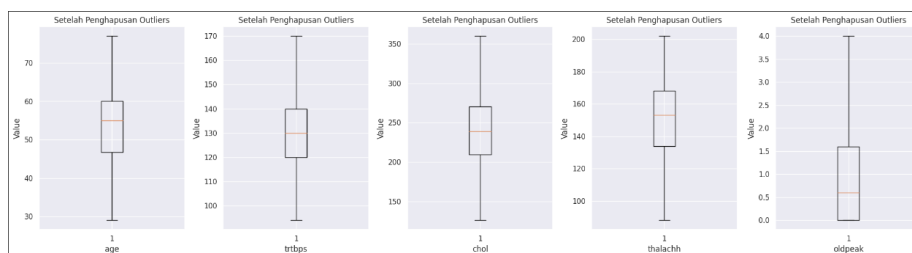
No	Model	Fitur	Accuracy	Precision	Recall	F1-Score
1	Model B	Kombinasi Fitur 4	0.91	0.94	0.89	0.91

#### 4.2.2 Pengujian Dengan Penghapusan *Outliers* dan Tanpa *SMOTE*

*Exploratory Data Analysis* (EDA) digunakan untuk memahami hubungan dari masing-masing variabel dengan kelas target, dan mencari *outliers* pada variabel numerical untuk mendapatkan hasil yang lebih optimal. Proses EDA dilakukan untuk variabel numerical dengan menggunakan *boxplot* untuk menunjukkan hubungan setiap variabel numerical dengan kelas target, serta untuk melihat adanya *outliers* pada setiap variabel numerical.

Gambar 4. 22 *Outliers* yang Terdeteksi

Pada Gambar 4.22 terdapat beberapa *outlier* yang terdeteksi pada proses EDA. Selanjutnya dilakukan penanganan *outlier* yang terdeteksi pada proses EDA tersebut dengan melakukan penghapusan *outliers* yang terdeteksi.

Gambar 4. 23 Setelah Penghapusan *Outliers*

Setelah dilakukan penanganan untuk *outlier* yang terdeteksi, jumlah data yang akan digunakan untuk pengujian adalah 284 data. Selanjutnya adalah membagi data latih dan data uji menjadi 3 model. Model A dengan perbandingan data latih (*train*) sebesar 90%, dan data uji (*test*) sebesar 10%. Model B dengan perbandingan data latih (*train*) sebesar 80%, dan data uji (*test*) sebesar 20%. Model C dengan perbandingan data latih (*train*) sebesar 75%, dan data uji (*test*) sebesar 25% (Sumwiza et al., 2023) dengan menggunakan 4 data kombinasi fitur yang berbeda.

Setelah pembagian data latih dan data uji menjadi 3 model dengan kombinasi 4 fitur yang berbeda, selanjutnya dilakukan *tuning hyperparameter* untuk setiap model yang ada dengan mengatur *n\_estimator* serta *max\_depth* untuk pengujian dengan penghapusan *outliers* – *SMOTE*. Hasil terbaik dari pengujian dengan penghapusan *outliers* – *SMOTE* tertera pada tabel berikut.

Tabel 4. 201 Hasil Terbaik Dari Pengujian Dengan Penghapusan Outliers Tanpa *SMOTE*

No	Model	Fitur	Accuracy	Precision	Recall	F1-Score
1	Model A	Kombinasi Fitur 3	0.90	0.93	0.87	0.90

Dari keempat pengujian yang telah dilakukan, hasil terbaik yang didapatkan dari masing-masing pengujian tertera pada tabel berikut.

Tabel 4. 202 Hasil Terbaik Dari Setiap Pengujian

No	Pengujian	Model	Fitur	Accuracy	Precision	Recall	F1-Score
1	SMOTE – Penghapusan Outliers	Model B	Kombinasi Fitur 2	0.84	0.87	0.82	0.84
2		Model B	Kombinasi Fitur 4	0.84	0.87	0.82	0.84
3	Penghapusan Outliers – SMOTE	Model A	Kombinasi Fitur 3	0.93	0.93	0.93	0.93
4		Model A	Kombinasi Fitur 4	0.93	0.93	0.93	0.93
5	Dengan SMOTE dan Tanpa Penghapusan Outliers	Model B	Kombinasi Fitur 4	0.91	0.94	0.89	0.91
6	Dengan Penghapusan Outliers dan Tanpa SMOTE	Model A	Kombinasi Fitur 3	0.90	0.93	0.87	0.90

Berdasarkan pada Tabel 4.202, dari keseluruhan pengujian yang telah dilakukan, pengujian dengan penghapusan *outliers – SMOTE* mendapatkan hasil yang terbaik untuk model A dengan kombinasi fitur 3 dan fitur 4 dengan nilai akurasi sebesar 93%. Model ini juga dapat dikatakan baik karena hasil evaluasi matriksnya dengan nilai akurasi, *precision*, *recall*, dan *f1-score* sama di angka 93%.

### 4.3 Integrasi Penelitian Dalam Tafsir Al-Qur'an

Penelitian ini dibangun untuk mengimplementasikan salah satu teknik dalam data *mining* yaitu klasifikasi, sebagaimana dijelaskan dalam Al-Qur'an Surah Al-An'am ayat 141:

وَهُوَ الَّذِي أَنْشَأَ جَنَّاتٍ مَّعْرُوشَاتٍ وَغَيْرَ مَعْرُوشَاتٍ وَالنَّخْلَ وَالزَّرْعَ مُخْتَلِفًا أَكْثَرًا ۗ وَالرَّيْحَانَ مُشَابِهًا وَغَيْرَ مُشَابِهٍ كُلُوا مِنْ ثَمَرِهِ إِذَا أَثْمَرَ وَآتُوا حَقَّهُ ۗ وَلَا تُسْرِفُوا إِنَّهُ لَا يُحِبُّ الْمُسْرِفِينَ

“Dan Dia-lah yang menjadikan tanaman-tanaman yang merambat dan yang tidak merambat, pohon kurma, tanaman beraneka ragam rasanya, zaitun dan delima yang serupa (bentuk dan warnanya) dan tidak serupa (rasanya). Makanlah

*buahnya apabila ia berbuah dan berikanlah haknya (zakatnya) pada waktu memetik hasilnya, tapi janganlah berlebih-lebihan. Sesungguhnya Allah tidak menyukai orang yang berlebih-lebihan.” (Q.S. Al-An’am: 141).*

Menurut tafsir Ibnu Katsir surah Al-An’am ayat 141 ini menjelaskan tentang ciptaan Allah SWT yang beragam termasuk tanaman. Allah SWT menciptakan tanaman dengan berbagai jenis, rasa, dan bentuk. Ayat ini juga mengingatkan manusia untuk bersyukur atas nikmat Allah SWT dengan cara memakan buah-buahan tersebut dan mengeluarkan zakatnya. Sama seperti Allah SWT yang menciptakan tanaman dengan berbagai jenis, rasa, dan bentuk, teknik klasifikasi dalam data *mining* digunakan untuk mengklasifikasikan data yang beragam berdasarkan pola dan kategorinya.

Penelitian ini dibangun dengan tujuan untuk mengklasifikasikan risiko serangan jantung, sebagai salah satu upaya untuk turut mencegah risiko serangan jantung, seperti yang dijelaskan dalam Al-Qur’an tentang pentingnya tolong-menolong dan berbuat kebaikan yaitu pada surah Al-Ma’idah ayat 2.

وَتَعَاوَنُوا عَلَى الْبِرِّ وَالتَّقْوَىٰ وَلَا تَعَاوَنُوا عَلَى الْإِثْمِ وَالْعُدْوَانِ يَوْمَئِذٍ اللَّهُ شَدِيدُ الْعِقَابِ

*“Tolong-menolonglah kamu dalam (mengerjakan) kebajikan dan takwa, dan jangan tolong menolong dalam berbuat dosa dan kemusuhan. Bertakwalah kepada Allah sesungguhnya Allah sangat berat siksaan-Nya” (Q.S. Al-Ma’idah: 2).*

Menurut Ibnu Katsir surah Al-Ma’idah ayat 2 menjelaskan perintah agar tolong-menolong dalam mengerjakan berbagai kebaikan dan tolong-menolong dalam menjauhi keburukan atau kemaksiatan (Tafsir Ibnu Katsir, n.d.). Seperti pada penelitian ini, implementasi ayat tersebut dengan memberikan pengetahuan bagaimana mengklasifikasikan populasi yang berisiko serangan jantung dengan

menggunakan metode *Random Forest* dengan harapan dapat mengetahui populasi yang berisiko serangan jantung maupun yang tidak berisiko serangan jantung. Selain itu untuk memberi informasi terkait faktor-faktor yang memicu risiko serangan jantung. Sistem yang telah dibangun diharapkan dapat membantu mengantisipasi sehingga dapat dilakukan penanganan yang sesuai. Selain pada Al-Qur'an, terdapat hadits yang berkaitan dengan penelitian ini, yaitu hadits riwayat Imam Muslim.

مَنْ نَفَسَ عَنْ مُؤْمِنٍ كُرْبَةً مِنَ الدُّنْيَا، نَفَسَ اللَّهُ عَنْهُ كُرْبَةً مِنْ كُرْبِ يَوْمِ الْقِيَامَةِ، وَمَنْ يَسَّرَ عَلَى مُعْسِرٍ، يَسَّرَ اللَّهُ عَلَيْهِ فِي الدُّنْيَا وَالْآخِرَةِ، وَمَنْ سَتَرَ مُسْلِمًا، سَتَرَهُ اللَّهُ فِي الدُّنْيَا وَالْآخِرَةِ، وَاللَّهُ فِي عَوْنِ الْعَبْدِ مَا كَانَ الْعَبْدُ فِي عَوْنِ أَخِيهِ

*“Barangsiapa melapangkan seorag mukmin dari satu kesusahan dunia, Allah akan melapangkannya salah satu kesusahan di hari kiamat. Barang siapa meringankan penderitaan seseorang, Allah akan meringankan penderitaannya di dunia dan akhirat, barang siapa menutupi (aib) seorang muslim, Allah aka menutupi (aib)nya di dunia dan akhirat. Allah akan menolong seorang hamba selama hamba itu mau menolong saudaranya.” (HR Muslim No 2699)*

Hadist ini mengajarkan tentang pentingnya sikap tolong-menolong dan membantu sesama, seperti yang diimplementasikan pada penelitian ini dalam hal pencegahan risiko serangan jantung. Saling membantu dan tolong-menolong merupakan salah satu prinsip fundamental dalam ajaran islam. Dengan menjalankan anjuran ini, umat muslim tidak hanya memperoleh pahala dan keberkahan dari Allah, tetapi juga turut berkontribusi untuk kehidupan bermasyarakat yang sejahtera.



## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berdasarkan skenario pengujian yang telah dilakukan sesuai dengan yang telah dijabarkan pada poin 3.8 dengan menggunakan 4 pengujian yaitu pengujian dengan *SMOTE* – penghapusan *outliers*, pengujian dengan penghapusan *outliers* – *SMOTE*, pengujian dengan *SMOTE* dan tanpa penghapusan *outliers*, dan pengujian dengan penghapusan *outliers* dan tanpa *SMOTE*. Pada masing-masing pengujian melibatkan 3 model, yaitu model A dengan rasio perbandingan data 90:10, model B dengan rasio perbandingan data 80:20, model A dengan rasio perbandingan data 75:25. Untuk setiap model dilakukan pengujian dengan kelompok data yang berdasarkan 4 kombinasi fitur yang diambil dari fitur penting data.

Hasil terbaik yang didapatkan dari total 48 pengujian yang telah dilakukan terdapat pada pengujian dengan penghapusan *outliers* – *SMOTE* pada model A dengan kombinasi fitur 3 dan kombinasi fitur 4 dengan nilai akurasi sebesar 93%. Model ini juga dapat dikatakan sangat baik karena hasil evaluasi matriksnya dengan nilai akurasi, *precision*, *recall*, dan *f1-score* sama di angka 93% (Subarkah et al., 2022). Selain itu, pada 16 pengujian yang telah dilakukan dengan menggunakan *data Real* atau keseluruhan data, algoritma *Random Forest* menunjukkan hasil yang sangat baik untuk setiap pengujiannya. Hasil yang didapatkan dari keseluruhan pengujian sesuai dengan data aktualnya, diketahui dari nilai akurasi, *precision*, *recall*, dan *f1 – score* yang menunjukkan nilai 100%.

Sehingga dapat disimpulkan bahwa algoritma *Random Forest* dapat mengklasifikasikan populasi yang berisiko serangan jantung berdasarkan data *Heart Attack Analysis & Prediction Dataset* dengan hasil yang baik.

## 5.2 Saran

Peneliti menyadari terdapat beberapa kekurangan dalam penelitian ini, sehingga diperlukannya kritik dan saran yang membangun agar dapat meningkatkan pada penelitian selanjutnya. Berdasarkan hasil kesimpulan yang telah diperoleh dari skenario pengujian, beberapa saran yang dapat diberikan untuk pengembangan dan penelitian selanjutnya adalah sebagai berikut:

1. Penggunaan dataset yang lebih beragam dari sumber yang berbeda agar dapat menilai kemampuan model dalam mengklasifikasikan pada populasi yang berbeda.
2. Pengujian dengan teknik *preprocessing* data yang lain yang mungkin dapat memberikan hasil yang lebih optimal dalam hal performa model.
3. Ekplorasi model dan algoritma lain agar dapat memberikan wawasan yang lebih luas mengenai keunggulan masing-masing metode.
4. Analisis lebih dalam terhadap kombinasi fitur lain yang mungkin dapat meningkatkan performa model.

## DAFTAR PUSTAKA

- Adi, S., & Wintarti, A. (2022). Komparasi Metode Support Vector Machine (Svm), K-Nearest Neighbors (Knn), Dan Random Forest (Rf) Untuk Prediksi Penyakit Gagal Jantung. *MATHunesa: Jurnal Ilmiah Matematika*, 10(2), 258–268. <https://doi.org/10.26740/mathunesa.v10n2.p258-268>
- Al Azhima, S. A. T., Darmawan, D., Arief Hakim, N. F., Kustiawan, I., Al Qibtiya, M., & Syafei, N. S. (2022). Hybrid Machine Learning Model untuk memprediksi Penyakit Jantung dengan Metode Logistic Regression dan Random Forest. *Jurnal Teknologi Terpadu*, 8(1), 40–46. <https://doi.org/10.54914/jtt.v8i1.539>
- Amrullah, S., Rosjidi, C., Dhessa, D., Wurjatmiko, A., & Hasrima. (2022). Faktor Resiko Penyakit Infark Miokard Akut di Rumah Sakit Umum Dewi Sartika Kota Kendari. *Jurnal Ilmiah Karya Kesehatan*, 02(02), 21–29. <https://stikesks-kendari.e-journal.id/JIKK/article/view/445/295>
- Annisa, R. (2019). Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung. *Jurnal Teknik Informatika Kaputama (JTIK)*, 3(1), 22–28. <https://jurnal.kaputama.ac.id/index.php/JTIK/article/view/141/156>
- Bhavani, A., & Santhosh Kumar, B. (2021). A Review of State Art of Text Classification Algorithms. *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021, Iccmc*, 1484–1490. <https://doi.org/10.1109/ICCMC51019.2021.9418262>
- Depari, D. H., Widiastiwi, Y., & Santoni, M. M. (2022). Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung. *Informatik: Jurnal Ilmu Komputer*, 18(3), 239. <https://doi.org/10.52958/iftk.v18i3.4694>
- Erdiansyah, U., Irmansyah Lubis, A., & Erwansyah, K. (2022). Komparasi Metode K-Nearest Neighbor dan Random Forest Dalam Prediksi Akurasi Klasifikasi Pengobatan Penyakit Kutil. *Jurnal Media Informatika Budidarma*, 6(1), 208. <https://doi.org/10.30865/mib.v6i1.3373>
- Gadjah Mada, U. (2021). *Menjaga Jantung Sejak Dini*. <https://ugm.ac.id/id/berita/21953-menjaga-jantung-sejak-dini-sangat-diperlukan/>
- Jin, Z., Shang, J., Zhu, Q., Ling, C., Xie, W., & Qiang, B. (2020). RFRSF: Employee Turnover Prediction Based on Random Forests and Survival

Analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12343 LNCS, 503–515. [https://doi.org/10.1007/978-3-030-62008-0\\_35](https://doi.org/10.1007/978-3-030-62008-0_35)

Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. *IOP Conference Series: Materials Science and Engineering*, 1022(1). <https://doi.org/10.1088/1757-899X/1022/1/012072>

Kadhim, M. A., & Radhi, A. M. (2023). Heart disease classification using optimized Machine learning algorithms. *Iraqi Journal for Computer Science and Mathematics*, 4(2), 31–42. <https://doi.org/10.52866/ijcsm.2023.02.02.004>

Kementerian Kesehatan. (2021). *Peringatan Hari Jantung Sedunia 2021: Jaga Jantungmu untuk Hidup Lebih Sehat*. <https://ayosehat.kemkes.go.id/peringatan-hari-jantung-sedunia-2021-jaga-jantungmu-untuk-hidup-lebih-sehat>

Kementerian Kesehatan. (2022). *Penyakit Jantung Penyebab Utama Kematian, Kemenkes Perkuat Layanan Primer*. Kemenkes.

Klusowski, J. M. (2018). Complete Analysis of a Random Forest Model. *ArXiv*, 13, 1063–1095.

Mount Elizabeth Hospital. (n.d.). *Infark Miokard (Serangan Jantung) - Gejala dan Penyebab*. <https://www.mountelizabeth.com.sg/id/conditions-diseases/myocardial-infarction/symptoms-causes>

Mourya, S., Srinivas, P. V. S., & Seetha, M. (2018). Multi-attributes web objects classification based on class-attribute relation patterns learning approach. *International Journal of Advanced Computer Science and Applications*, 9(12), 409–416. <https://doi.org/10.14569/IJACSA.2018.091258>

Mukhiya, S. K., & Ahmed, U. (2020). *Hands-On Exploratory Data Analysis with Python: Perform EDA techniques to understand, summarize, and investigate your data*. Packt Publishing Ltd.

Niranjan, O., & Dhamoon, A. S. (2023). *Myocardial Infarction*. National Library of Medicine. <https://www.ncbi.nlm.nih.gov/books/NBK537076/>

Nugraha, W. (2021). Prediksi Penyakit Jantung Cardiovascular Menggunakan Model Algoritma Klasifikasi. *Jurnal Manajemen Dan Informatika*, 9(2), 3–8.

Qomariyah, N., Hamzah, N., Prima Mustika, W., Studi Sistem Informasi, P., &

- Nusa Mandiri, S. (2018). Inti Nusa Mandiri Penerapan Metode Artificial Neural Network Untuk Mendeteksi Serangan Jantung Di Rs Awal Bros Bekasi. *Inti Nusa Mandiri*, 13(1), 27–32. [www.nusamandiri.ac.id](http://www.nusamandiri.ac.id)
- Radhika, R., & Thomas George, S. (2021). Heart Disease Classification Using Machine Learning Techniques. *Journal of Physics: Conference Series*, 1937(1), 1137–1144. <https://doi.org/10.1088/1742-6596/1937/1/012047>
- Rohayati, & Widani, N. L. (2020). Analisis Faktor yang Berhubungan dengan Serangan Jantung Berulang pada Pasien Post Percutaneous Transluminal Coronary Angioplasty. *Jurnal Kesehatan*, 10, 25–39.
- Sahoo, K., Samal, A. K., Pramanik, J., & Pani, S. K. (2019). Exploratory data analysis using python. *International Journal of Innovative Technology and Exploring Engineering*, 8(12), 4727–4735. <https://doi.org/10.35940/ijitee.L3591.1081219>
- Scarano, A., Rella Riccardi, M., Mauriello, F., D’Agostino, C., Pasquino, N., & Montella, A. (2023). Injury severity prediction of cyclist crashes using random forests and random parameters logit models. *Accident Analysis & Prevention*, 192(September), 107275. <https://doi.org/10.1016/j.aap.2023.107275>
- Subarkah, P., Damayanti, W. R., & Permana, R. A. (2022). Comparison of Correlated Algorithm Accuracy Naive Bayes Classifier and Naive Bayes Classifier for Classification of heart failure. *ILKOM Jurnal Ilmiah*, 14(2), 120–125. <https://doi.org/10.33096/ilkom.v14i2.1148.120-125>
- Sumwiza, K., Twizere, C., Rushingabigwi, G., Bakunzibake, P., & Bamurigire, P. (2023). Enhanced cardiovascular disease prediction model using random forest algorithm. *Informatics in Medicine Unlocked*, 41(July), 101316. <https://doi.org/10.1016/j.imu.2023.101316>
- Tafsir Ibnu Katsir. (n.d.). *Tafsir Ibnu Katsir Online*. <http://www.ibnukatsironline.com/2015/10/tafsir-surat-al-haqqah-ayat-44-52.html>