Jurnal Teknologi dan Manajemen Industri Terapan (JTMIT) Vol. 4, No. 4, Desember 2025 pp. 1230 - 1237
P-ISSN: 2829-0232 E-ISSN: 2829-0038

Klasifikasi Penyakit Jantung dengan Metode Random Forest

Ahmad Muflih Wafir S.A[1], Ahmad Homaidi[2], Hermanto[3]
[1,2,3)] Jurusan Teknologi Informasi, Fakultas Sains dan Teknnologi, Universitas Ibrahimy Situbondo
Jl. KHR. Syamsul Arifin No.1-2, Sukorejo, Sumberejo, Kec. Banyuputih, Kabupaten Situbondo, Jawa Timur 68374
Email: muflih438@gmail.com , ahmadhomaidi@ibrahimy.ac.id , otnamreh54@gmail.com

ABSTRAK

Penyakit jantung merupakan salah satu penyebab kematian tertinggi di dunia dengan angka kematian mencapai 17,8 juta jiwa setiap tahun menurut World Health Organization (WHO). Di Indonesia, prevalensi penyakit jantung terus meningkat dari tahun ke tahun, dipengaruhi oleh faktor seperti pola makan tidak sehat, kurangnya aktivitas fisik, dan faktor genetik. Deteksi dini menjadi kunci penting untuk mencegah komplikasi serius dan meningkatkan peluang kesembuhan. Perkembangan teknologi informasi, khususnya di bidang machine learning, membuka peluang untuk membantu proses diagnosis penyakit jantung. Salah satu algoritma yang banyak digunakan dan terbukti efektif adalah Random Forest, yang termasuk dalam kategori ensemble learning dan mampu meminimalkan risiko overfitting serta menangani data berskala besar dan non-linear.Penelitian ini menggunakan dataset UCI Heart Disease dengan tahapan pra-pemrosesan meliputi penghapusan outlier, penyeimbangan data menggunakan Synthetic Minority Oversampling Technique (SMOTE), serta normalisasi menggunakan StandardScaler. Pengujian dilakukan 80:20 dengan empat kombinasi fitur berdasarkan nilai kepentingan dari atribut feature_importances_. Hasil pengujian menunjukkan bahwa kombinasi penghapusan outlier diikuti SMOTE memberikan performa terbaik dengan nilai AUC tertinggi 0,96, diikuti penghapusan outlier saja dengan AUC 0,95, dan SMOTE saja dengan AUC 0,94.

Kata kunci: Random Forest, Penyakit Jantung, Deteksi Dini, Klasifikasi, Resiko, Machine Learning.

ABSTRACT

Heart disease is one of the leading causes of death worldwide, with an estimated 17.8 million deaths each year according to the World Health Organization (WHO). In Indonesia, the prevalence of heart disease continues to increase annually, influenced by factors such as unhealthy dietary habits, lack of physical activity, and genetic predisposition. Early detection is crucial in preventing serious complications and improving recovery prospects. The advancement of information technology, particularly in machine learning, offers opportunities to support the diagnostic process of heart disease. One widely used and proven effective algorithm is Random Forest, which belongs to the ensemble learning category and is capable of minimizing the risk of overfitting while handling large-scale and non-linear data. This study utilises the UCI Heart Disease dataset with preprocessing stages including outlier removal, data balancing using the Synthetic Minority Oversampling Technique (SMOTE), and normalisation using StandardScaler. The testing used an 80:20 ratio with four feature combinations based on the importance values from the feature_importances_ attribute. The results indicate that the combination of outlier removal followed by SMOTE yielded the best performance, achieving the highest AUC score of 0.96, followed by outlier removal alone with an AUC of 0.95, and SMOTE alone with an AUC of 0.94.

Keywords: *Random Forest, Heart Disease, Early Detection, Classification, Risk, Machine Learning.*

Pendahuluan

Penyakit jantung merupakan salah satu penyebab kematian tertinggi di dunia. Menurut data World Health Organization (WHO), penyakit kardiovaskular menyumbang sekitar 17,8 juta kematian setiap tahun. Di Indonesia, prevalensi penyakit jantung mengalami peningkatan signifikan dari tahun ke tahun, untuk penyakit jantung mencapai 13% dari total kematian di dunia. Di Indonesia, yang terjangkit penyakit jantung ini sangat meningkat yang signifikat dari tahun ke tahun. Banyak faktor yang penyebabkan hal itu terjadi seperti pola makan tidak sehat, kurangnya aktivitas visik atau faktor genetik dari keluarga. Deteksi dini penyakit jantung sangatlah penting untuk mencegah komplikasi serius dan meningkatkan peluang kesembuhan.[1]

Pencegahan sejak dini pada risiko penyakit serangan jantung haruslah dengna kesadaran setiap peribadi yang sedang terjangkit untuk pentingnya menjaga pola hidup yang sehat, serta harus menjahui pemicu terjadinya serangan jantung yang sangat berisiko menyebabkan kematian, tentunya diperlukan pemahaman yang lebih mendalam terkait faktor-faktor resiko dan identifikasi terhadap indivdu yang memiliki resiko tinggi.[2]

Dengan seiring perkembangan teknologi informasi, pengguna komputasi cerdas khusunya dalam bidang Machine Learning menjadi salah satu solusi yang berpotensi dalam membantu proses diagnosis penyakit. Salah satu metode yang sering digunakan dalam mengklasifikasi dalam dunia medis adalah Random Forest. [3]

Metode Random Forest adalah Algoritma yang termasuk dalam kategori Ensamble Learning dan salah satu Algoritma pengajaran mesin yang terkenal. Ensamble Learning melibatkan penggambungan hasil dari beberapa model untuk meningkatkan kinerja dan ketetapan prediksi dibandingkan dengan penggunaan satu model tunggal.[4]

Berdasarkan dari peneliti sebelumnya yang dilakukan oleh Tony Raymond dan Allan Desi Alexander pada tahu 2024 bertujuan memprediksi resiko serangan jantung dengan Algoritma Random Forest. Data yang digunakan dalam analisis ini dengan dataset yang berisi 303 Sampel dan 14 atribut diambil dari data sekunder yang digunakan diproleh dari website Kaggle. Hasil eksperimen menunjukan dengan acak Algoritma forest, yang diproleh akurasi sebesar 86,9% untuk predeksi penyakit jantung dengan snilai sensitivitas 90,6%, dan nilai spesifikasi 82,7%.[5]

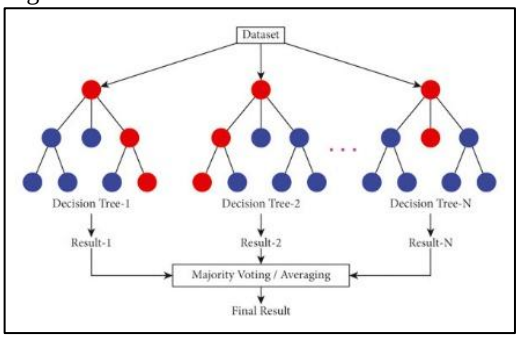
Dengan perkembangan teknologi yang pesat, kini ada solusi yang cukup menjanjikan, yaitu penggunaan machine learning dalam dunia medis. Salah satu algoritma yang sudah terbukti efektif dalam melakukan klasifikasi atau prediksi adalah Random Forest.

Metode Penelitian

Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimen semu (quasi-experimental). Data yang digunakan merupakan data sekunder yang diperoleh dari situs Kaggle, yaitu dataset yang memuat informasi terkait karakteristik pasien seperti usia, jenis kelamin, tekanan darah, kadar kolesterol, detak jantung maksimal, dan variabel medis lainnya yang berhubungan dengan penyakit jantung.[6]

Algoritma

Pada penelitian ini akan digunakan metode random forest, Algoritma ini bekerja dengan membangun banyak pohon keputusan (decision trees) dan menggabungkan hasilnya melalui proses voting untuk menghasilkan prediksi yang lebih akurat. Random Forest unggul dalam mengatasi overfitting dan bekerja efektif baik pada data klasifikasi maupun regresi[7] Algoritma ini membuat klasifikasi dengan manggambungkan klasifikasi dari semua pohon keputusan dimana pohon yang dinilai terbaik oleh mesin yang akan digunakan.



Gambar 1. Struktur Random Forest.

Alur dasar dari algoritma Random Forest seperti pada gambar dimulai dengan data training dimuat kedalam memori, kemudian dibagi menjadi subset acak menggunakan teknik boostrap. Teknik boostrap 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 kedalam pohon atau maksimum fitur yang dapat digunakan untuk setiap split untuk menghindari overfitting.[8], [9], [10], [11], [12], [13]

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. [14]

Untuk langkah-langkah yang dilakukan oleh peneliti dalam menerapkan klasifikasi random forest, sebagai berikut :

Pengambilan dataset

Dalam proses awal ini, penulis mencari dataset yang sesuai dengan topik penelitian ini, sehingga penulis mengambil dataset publik dari Kaggle yang nantinya akan digunakan dalam proses pengolahan klasifikasi.[15]

Split Data

Tahap ini melibatkan pemisahan dataset menjadi data pelatihan dan data pengujian. Dalam studi ini, pembagiannya adalah 80/20. Umumnya, model pembelajaran mesin mencapai hasil akurasi yang baik ketika

berisi sejumlah kecil data pengujian. Oleh karena itu, dalam studi ini, kami menambah data pengujian dan menguji apakah model tersebut mencapai hasil yang baik.[16]

Klasifikasi dengan Random Forest

Langkah berikutnya adalah klasifikasi menggunakan metode Random Forest, pada titik ini ditentukan apakah pasien diketahui menderita penyakit atau tidak.[17]

Confusioon Matrix

Confusion Matrix adalah satu istilah mendasar dalam machine learning. Confusion Matrix digunakan untuk mengukur akurasi model dengan cara membandingkan nilai prediksi dan juga nilai aktual [18]. Berikut ini tabel Confusion Matrix menurut [19].

Tabel 1. Confusion Matrix

Kelas	**Aktual Positif**	**Negatif**
Prediksi Positif	Positif Benar *TP*	Positif Palsu FP
Negatif	Negatif Palsu *FB*	Negatif Benar TN

Evaluasi Model

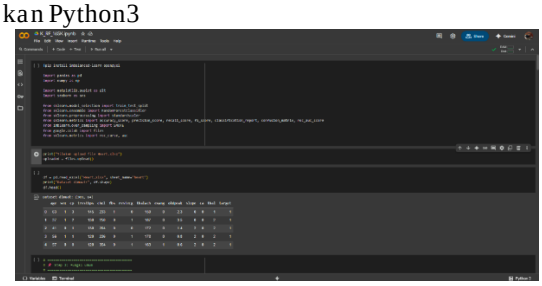
Untuk menilai performa model klasifikasi, digunakan beberapa metrik evaluasi seperti akurasi, precision, recall, dan F1-score, akurasi mengukur seberapa banyak prediksi yang benar dari seluruh prediksi yang dilakukan,[20] Precision adalah proporsi data positif yang diprediksi benar terhadap seluruh data yang diprediksi positif, Recall menunjukkan seberapa banyak data positif yang berhasil dikenali oleh model, F1-score adalah rata-rata harmonis dari precision dan recall, dan digunakan saat distribusi kelas tidak seimbang[21], [22], [23], [24], [25]

Hasil Uji Coba

Poin ini menjabarkan analisis hasil pengujian berdasarkan skenario pengujian 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

Pengujian Software yang Digunakan

Pengujian terhadap penggunaan Google Colab sebagai text editor untuk membuat program klasifikasi dengan bahasa pemrograman Python3 menunjukkan kinerja yang efektif dan efisien. Platform ini dilengkapi dengan fitur terintegrasi, mendukung beragam library Python secara bawaan, serta memungkinkan eksekusi kode secara interaktif. Fasilitas seperti auto-completion, syntax highlighting, dan integrasi dengan Google Drive membantu meningkatkan produktivitas dalam penulisan serta penyimpanan kode. Namun, pengguna memerlukan koneksi internet yang stabil dan harus mempertimbangkan keterbatasan akses terhadap file sistem lokal. Secara keseluruhan, Google Colab terbukti menjadi sarana yang efektif dan efisien untuk pengembangan program klasifikasi menggunakan Python3



Gambar 2. Hasil Pengujian Software

Model Random Forest

mencakup beberapa operasi pembuatan model RF, seperti menentukan nilai untuk parameter "n_estimators", "max_features", dan "min_samples_leaf". Kemudian, ditentukan label prediktor dan model disesuaikan dengan kumpulan data gabungan, yaitu data pelatihan dan data uji. Untuk klasifikasi, model yang dibangun menggunakan beberapa parameter akan diprediksi berdasarkan data uji terpisah.[26]

Pengujian

Dalam pengujian ini, dilakukan penghapusan outlier yang terdeteksi, kemudian diterapkan teknik Synthetic Minority Oversampling Technique (SMOTE) untuk mengatasi ketidakseimbangan data. Setelah penerapan SMOTE dan penghapusan outlier, jumlah data yang digunakan menjadi 318 entri. Pengujian dilakukan dengan rasio 80:20.

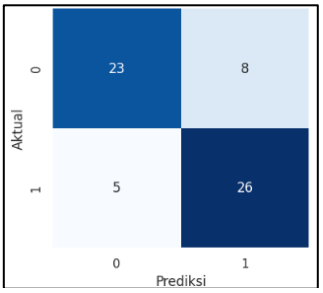
Setelah model dilatih, tingkat kepentingan setiap fitur diperoleh melalui atribut feature_importances_, yang menunjukkan kontribusi masing-masing fitur dalam proses pembagian node pada pohon keputusan. Nilai kepentingan dari beberapa model yang dilatih pada subset data berbeda kemudian dirata-ratakan untuk memperoleh estimasi yang lebih stabil. Selanjutnya, setiap model diuji menggunakan empat kombinasi fitur yang berbeda, berdasarkan fitur-fitur penting yang telah ditentukan sebelumnya.

Pemodelan Klasifikasi

Pada model yang di pakai 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, bootsrap = true, dan max_depth = 50. Pada pengujian dilakukan 4 kali pengujian dengan 4 kombinasi fitur yang berbeda.

Kombinasi fitur pertama

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



Gambar 3. Confusion Matrix Skenario

Tabel 2. Confusion Matrix Pengujian Model B Kombinasi Fitur Pertama

TP	TN	FP	FN
23	26	8	5

Berdasarkan Tabel 2, model Random Forest Classifier dalam proses klasifikasi menghasilkan 23 data dengan prediksi 0 (tidak berisiko serangan jantung) yang sesuai dengan hasil aktual, serta 26 data dengan prediksi 1 (berisiko serangan jantung) yang juga sesuai dengan hasil aktual. Namun, terdapat 8 data dengan prediksi 0 (tidak berisiko serangan jantung) tetapi hasil aktualnya adalah 1 (berisiko serangan jantung), dan 5 data dengan prediksi 1 (berisiko serangan jantung) namun hasil aktualnya 0 (tidak berisiko serangan jantung). Hasil matriks evaluasi untuk pengujian ini disajikan sebagai berikut.

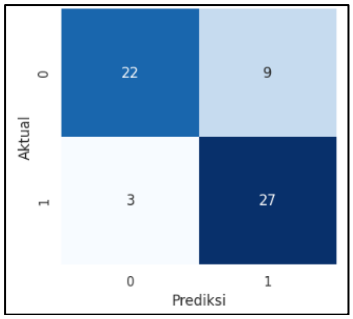
$$Akurasi = \frac{23 + 26}{23 + 26 + 8 + 5} = 0.790323$$

Tabel 3. Hasil Evaluasi Pengujian Model B Kombinasi Fitur Pertama

Accuracy	*Precision*	*Recall*	*F1-Score*
0.790323	*0.764706*	*0.838710*	*0.800000*

Kombinasi fitur Kedua

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



Gambar 4. Confusion Matrix Skenario 6

Tabel 4. Confusion Matrix Pengujian Model B Kombinasi Fitur Kedua

TP	TN	FP	FN
22	27	9	3

Berdasarkan Tabel 4, model Random Forest Classifier dalam melakukan klasifikasi menghasilkan 22 data dengan prediksi 0 (tidak berisiko serangan jantung) yang sesuai dengan hasil aktual, serta 27 data dengan prediksi 1 (berisiko serangan jantung) yang juga sesuai dengan hasil aktual. Sementara itu, terdapat 3 data dengan prediksi 0 (tidak berisiko serangan jantung) namun hasil aktualnya adalah 1 (berisiko serangan jantung), dan 9 data dengan prediksi 1 (berisiko serangan jantung) tetapi hasil aktualnya adalah 0 (tidak berisiko serangan jantung). Hasil matriks evaluasi untuk pengujian ini disajikan sebagai berikut.

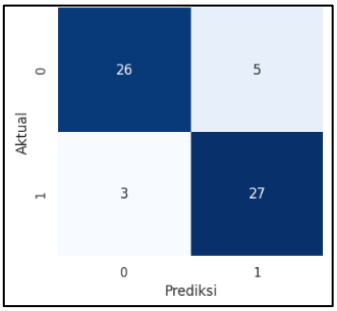
$$Akurasi = \frac{22 + 27}{22 + 27 + 9 + 3} = 0.803279$$

Tabel 5. Hasil Evaluasi Pengujian Model B Kombinasi Fitur Kedua

Accuracy	*Precision*	*Recall*	*F1-Score*
0.803279	*0.750000*	*0.900000*	*0.818182*

Kombinasi fitur Ketiga

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



Gambar 5. Confusion Matrix Skenario

Tabel 6. Confusion Matrix Pengujian Model B Kombinasi Fitur Ketiga

TP	TN	FP	FN
26	27	3	5

Berdasarkan Tabel 5, model Random Forest Classifier dalam proses klasifikasi menghasilkan 22 data dengan prediksi 0 (tidak berisiko serangan jantung) yang sesuai dengan hasil aktual, serta 25 data dengan prediksi 1 (berisiko serangan jantung) yang juga sesuai dengan hasil aktual. Selain itu, terdapat 5 data dengan prediksi 0 (tidak berisiko serangan jantung) namun hasil aktualnya adalah 1 (berisiko serangan jantung), dan 5 data dengan prediksi 1 (berisiko serangan jantung) tetapi hasil aktualnya adalah 0 (tidak berisiko serangan jantung). Hasil matriks evaluasi untuk pengujian ini disajikan sebagai berikut.

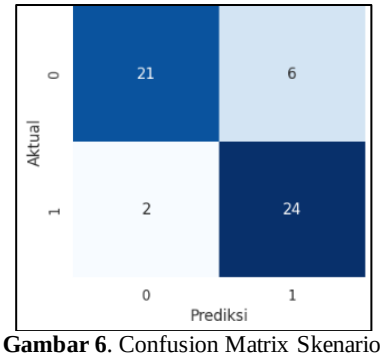
$$Akurasi = \frac{26 + 27}{26 + 27 + 5 + 3} = 0.868852$$

Tabel 7. Hasil Evaluasi Pengujian Model B Kombinasi Fitur Ketiga

Accuracy	*Precision*	*Recall*	*F1-Score*
0.868852	*0.843750*	*0.900000*	*0.87096*

Kombinasi Semua Fitur

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



Gambar 6. Confusion Matrix Skenario

Tabel 8. Confusion Matrix Pengujian Model B Kombinasi Semua Fitur

TP	TN	FP	FN
21	24	6	2

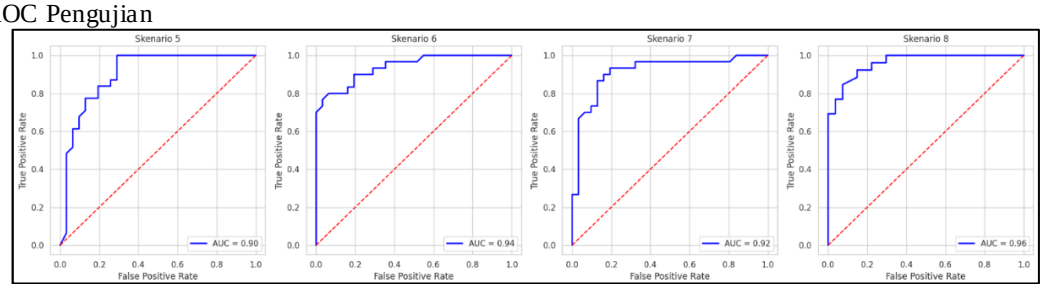
Berdasarkan Tabel 8, model Random Forest Classifier dalam melakukan klasifikasi menghasilkan 21 data dengan prediksi 0 (tidak berisiko serangan jantung) yang sesuai dengan hasil aktual, serta 24 data dengan prediksi 1 (berisiko serangan jantung) yang juga sesuai dengan hasil aktual. Sementara itu, terdapat 2 data dengan prediksi 0 (tidak berisiko serangan jantung) namun hasil aktualnya adalah 1 (berisiko serangan jantung), dan 6 data dengan prediksi 1 (berisiko serangan jantung) tetapi hasil aktualnya adalah 0 (tidak berisiko serangan jantung). Hasil matriks evaluasi untuk pengujian ini disajikan sebagai berikut.

$$Akurasi = \frac{21 + 24}{21 + 24 + 2 + 6} = 0.849057$$

Tabel 9. Hasil Evaluasi Pengujian Model B Kombinasi Semua Fitur

Accuracy	***Precision***	***Recall***	***F1-Score***
0.849057	*0.800000*	*0.923077*	*0.857143*

Hasil ROC Pengujian



Gambar 7. ROC Curve untuk Skenario 1 hingga 4

Gambar berikut menampilkan *ROC Curve* untuk Skenario 1 hingga 4, di mana tahap *preprocessing* dilakukan dengan menghapus outlier terlebih dahulu, kemudian menerapkan *SMOTE* untuk penyeimbangan kelas.

Skenario 1 – Memperoleh AUC sebesar 0,90, menunjukkan kemampuan diskriminasi model yang baik dalam membedakan kelas positif dan negatif. Bentuk kurva yang menonjol ke kiri atas mengindikasikan sensitivitas yang cukup tinggi dengan tingkat kesalahan prediksi positif yang rendah.

Skenario 2 – Nilai AUC meningkat menjadi 0,94, menandakan perbaikan signifikan pada kinerja model. Hal ini menunjukkan bahwa penghapusan outlier memberikan data yang lebih bersih sehingga *SMOTE* dapat menghasilkan sampel sintetis yang lebih representatif.

Skenario 3 – Mendapatkan AUC 0,92, sedikit lebih rendah dibandingkan Skenario 6 namun tetap menunjukkan performa yang sangat baik. *ROC Curve* tetap melengkung tajam ke arah kiri atas, yang berarti model seimbang dalam hal sensitivitas dan spesifisitas.

Skenario 4 – Mencapai AUC tertinggi dalam kelompok ini, yaitu 0,96. Nilai ini mengindikasikan kemampuan deteksi yang sangat tinggi dengan tingkat kesalahan prediksi yang sangat kecil.

Secara keseluruhan, pola yang terlihat pada kelompok ini menunjukkan bahwa penghapusan outlier sebelum penerapan *SMOTE* dapat menghasilkan data yang lebih optimal untuk pelatihan model, sehingga nilai AUC cenderung tinggi dan stabil.

Simpulan

Penelitian ini memanfaatkan metode Random Forest Classifier untuk mengklasifikasikan tingkat risiko penyakit jantung dengan dukungan tahap praproses data yang mencakup penghapusan outlier dan penerapan Synthetic Minority Oversampling Technique (SMOTE). Langkah ini bertujuan untuk mengatasi permasalahan ketidakseimbangan kelas sekaligus memastikan kualitas data yang lebih representatif sebelum proses pelatihan model. Dataset dibagi menggunakan rasio 80% data latih dan 20% data uji, kemudian diuji menggunakan beberapa kombinasi fitur yang dipilih berdasarkan tingkat kepentingan yang dihasilkan dari model.

Berdasarkan hasil pengujian, Kombinasi Fitur 3 yang memanfaatkan 11 fitur teratas menunjukkan performa paling optimal dengan akurasi 86,88%, precision 84,38%, recall 90%, serta F1-score 87,10%. Sementara itu, pengujian menggunakan seluruh fitur pada Skenario 4 menghasilkan nilai Area Under Curve (AUC) tertinggi sebesar 0,96, yang menunjukkan kemampuan model dalam membedakan kelas positif dan negatif secara sangat baik. Secara umum, seluruh skenario menghasilkan akurasi di atas 79%, menandakan bahwa model memiliki konsistensi performa yang baik.

Temuan ini memperlihatkan bahwa penerapan penghapusan outlier sebelum SMOTE mampu menghasilkan data yang lebih optimal untuk pelatihan, sehingga berdampak positif pada stabilitas dan kinerja model. Dengan akurasi yang tinggi serta nilai AUC yang signifikan, model Random Forest yang dikembangkan pada penelitian ini memiliki potensi untuk diimplementasikan sebagai sistem pendukung keputusan di bidang kesehatan, khususnya dalam deteksi dini risiko penyakit jantung. Penerapan sistem semacam ini diharapkan dapat membantu tenaga medis dalam melakukan pencegahan dan penanganan lebih cepat, sehingga dapat meminimalkan risiko komplikasi serius pada pasien.

Daftar Pustaka

[1] "Penyakit-Jantung-Koroner-Akut-Serangan-Jantung.Pdf."

[2] A.Putra, Agung Sagung Mas Meiswaryasti *Et Al.*, "Edukasi Penangan Awal Pada Serangan Jantung," *J. Pengabdi. Magister Pendidik. IPA*, Vol. 6, No. 4, Pp. 4–7, 2023.

[3] A. Y.Agusyul Andf.Firmansyah, "Prediksi Penyakit Jantung Menggunakan Algoritma Random Forest," *J. Minfo Polgan*, Vol. 12, No. 2, Pp. 2239–2246, 2023, Doi: 10.33395/Jmp.V12i2.13214.

[4] T.Informatika, "1 Teknik Informatika 2019," Pp. 1–7, 2019.

[5] T.Nangon, "Prediksi Tahap Awal Penyakit Jantung Menggunakan Algoritma Random Forest (Studi Kasus RSIJ)," *Da'watuna J. Commun. Islam. Broadcast*, Vol. 4, No. 4, Pp. 1561–1567, 2024, Doi: 10.47467/Dawatuna.V4i4.1882.

[6] N. H.Alfajr Ands.Defiyanti, "Metode Random Forest Dan Penerapan Principal Component Analysis (PCA)," Vol. 12, No. 3, 2024.

[7] M. K.Dwipa Jaya, "Perbandingan Random Forest, Decision Tree, Gradient Boosting, Logistic Regression Untuk Klasifikasi Penyakit Jantung," *Jnatia*, Vol. 2, No. November, Pp. 1–5, 2023.

[8] A.Ghozali, H.Pratiwi, and. S.Handajani, "Implementasi Data Mining Menggunakan Metode Random Forest Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes," *Delta J. Ilm. Pendidik. Mat.*, Vol. 11, No. 2, P. 147, 2023, Doi: 10.31941/Delta.V11i2.2686.

[9] M.Rizki, D.Devrika, Andi. H.Umam, "Aplikasi Data Mining Dalam Penentuan Layout Swalayan Dengan Menggunakan Metode MBA," *J. Tek. Ind. J. Has. Penelit. Dan Karya Ilm. Dalam Bid. Tek. Ind.*, Vol. 5, No. 2, Pp. 130–138, 2020.

[10] H.Chen, K. P.Xu, L. F.Chen, Andq. S.Jiang, "Self-Expressive Kernel Subspace Clustering Algorithm For Categorical Data With Embedded Feature Selection," *Mathematics*, Vol. 9, No. 14, 2021, Doi: 10.3390/Math9141680.

[11] R. J.Kuo, Y. R.Zheng, Andt. P. Q.Nguyen, "Metaheuristic-Based Possibilistic Fuzzy *K*-Modes Algorithms For Categorical Data Clustering," *Inf. Sci. (Ny).*, Vol. 557, Pp. 1–15, 2021, Doi: 10.1016/J.Ins.2020.12.051.

[12] J.Martinez, "Customer-Centric Framework: The Missing Link Between Data And Business Value," *Appl. Mark. Anal.*, Vol. 7, No. 4, Pp. 345–354, 2022, [Online]. Available: Https://Api.Elsevier.Com/Content/Abstract/Scopus_Id/85127620919

[13] Z.Xia, "Named Entity Recognition Of HSE Inspection Minutes Based On Data Enhancement," *China Saf. Sci. J.*, Vol. 32, No. 12, Pp. 53–62, 2022, Doi: 10.16265/J.Cnki.Issn1003-3033.2022.12.2727.

[14] F. S.Wijaya, M. F.Arotsid, Andr.Adriano, "Literatur Review : Penerapan Algoritma Random Forest Untuk Klasifikasi Penyakit Diabetes," Vol. 2, No. 3, Pp. 474–481, 2024.

[15] K.Widya Kayohana, "Klasifikasi Penyakit Hati Menggunakan Random Forest Dan Knn," *JATI (Jurnal Mhs. Tek. Inform.*, Vol. 8, No. 4, Pp. 7924–7929, 2024, Doi: 10.36040/Jati.V8i4.10457.

[16] S. A.Putri, N.Selayanti, Andm.Kristanaya, "Penerapan Machine Learning Algoritma Random Forest Untuk Prediksi Penyakit Jantung," Vol. 2024, No. Senada, 2024.

[17] H.Hidayat, A.Sunyoto, Andh.Alfatta, "Klasifikasi Penyakit Jantung Menggunakan Random Forest

Clasifier," *J. SISKOM-KB (Sistem Komput. Dan Kecerdasan Buatan)*, Vol. 7, No. 1, Pp. 31–40, 2023, Doi: 10.47970/Siskom-Kb.V7i1.464.

[18] M.Rizki *Et Al*., "Aplikasi End User Computing Satifaction Pada Penggunaan E-Learning FST UIN SUSKA," *SITEKIN J. Sains, Teknol. Dan Ind*., Vol. 19, No. 2, Pp. 154–159, 2022, Accessed: Jun.05, 2022. [Online]. Available: Http://Ejournal.Uin-Suska.Ac.Id/Index.Php/Sitekin/Article/View/14730

[19] W.Tanujaya, D.Dewi, D. E.-W.Teknik, And Undefined2013, "Penerapan Algoritma Genetik Untuk Penyelesaian Masalah Vehicle Routing Di PT. MIF," *Journal.Wima.Ac.Id*, Accessed: Feb.08, 2022. [Online]. Available: Http://Journal.Wima.Ac.Id/Index.Php/Teknik/Article/View/163

[20] R. F. N.Iskandar, D. H.Gutama, D. P.Wijaya, Andd.Danianti, "Klasifikasi Menggunakan Metode Random Forest Untuk Awal Deteksi Diabetes Melitus Tipe 2," *J. Tek. Ind. Terintegrasi*, Vol. 7, No. 3, Pp. 1620–1626, 2024, Doi: 10.31004/Jutin.V7i3.26916.

[21] D. H.Depari, Y.Widiastiwi, Andm. M.Santoni, "Perbandingan Model Decision Tree, Naive Bayes Dan Random Forest Untuk Prediksi Klasifikasi Penyakit Jantung," *Inform. J. Ilmu Komput*., Vol. 18, No. 3, P. 239, 2022, Doi: 10.52958/Iftk.V18i3.4694.

[22] S.Hu, "A Data-Centric Solution To Improve Online Performance Of Customer Service Bots," 2023. Doi: 10.1145/3539618.3591843.

[23] D.Markudova, "Preventive Maintenance For Heterogeneous Industrial Vehicles With Incomplete Usage Data," *Comput. Ind..*, Vol. 130, 2021, Doi: 10.1016/J.Compind.2021.103468.

[24] R.Shirkhorshidi, "How To Comply With Generated Data With Wells Hse Performance Procedures And Requirements In Practice," 2023. Doi: 10.2118/214598-MS.

[25] T.Karabchuk, "Motherhood Wage Penalty In Russia: Empirical Study On RLMS-HSE Data," 2021. [Online]. Available: Https://Api.Elsevier.Com/Content/Abstract/Scopus_Id/85138947354

[26] A. M.Alfaudzan, A.Almayda, Andr.Nugraha, "Perbandingan Model Data Mining Klasifikasi Dalam Memprediksi Penyakit Jantung Di Eropa," 2022.