JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9485
Hal 398-409
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Studi Komparatif Model Machine Learning untuk Klasifikasi Penyakit Jantung dengan SMOTE pada Data Imbalanced

Glen Fierre Sijabat*, Wahyu Aji Eko Prabowo

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1*]111202214369@mhs.dinus.ac.id, [2]prabowo@dsn.dinus.ac.id
Email Penulis Korespondensi: 111202214369@mhs.dinus.ac.id
Submitted **07-01-2026;** Accepted **24-02-2026;** Published **28-02-2026**

Abstrak

Penelitian ini membahas penerapan Synthetic Minority Over-sampling Technique (SMOTE) pada klasifikasi penyakit jantung menggunakan empat algoritma pembelajaran mesin, yaitu Logistic Regression, Random Forest, LightGBM, dan XGBoost, berdasarkan data Heart Disease UCI yang berisi 920 rekam medis dengan 16 fitur klinis. Label awal tingkat keparahan (0-4) dikonversi menjadi dua kelas, yaitu tidak sakit (0) dan sakit (1-4), sehingga lebih sesuai untuk kebutuhan skrining klinis berbasis keputusan biner. Eksperimen dilakukan dalam dua skenario: (1) pelatihan model pada data asli tanpa penanganan ketidakseimbangan kelas dan (2) pelatihan model dengan SMOTE yang hanya diterapkan pada data latih di dalam pipeline, disertai proses hyperparameter tuning menggunakan k-fold cross-validation. Kinerja model dievaluasi dengan metrik akurasi, presisi, recall, F1-score, AUC-ROC, serta analisis confusion matrix untuk mengkaji kesalahan klasifikasi, khususnya kasus false negative pada kelas pasien sakit. Pada skenario tanpa SMOTE, model terbaik Logistic Regression mencapai akurasi 84,78%, recall 84,31%, F1-score 86,00%, dan AUC-ROC 91,95%, namun jumlah false negative masih relatif tinggi. Setelah penerapan SMOTE, terjadi peningkatan recall dan F1-score kelas positif pada semua model, dengan kinerja terbaik diperoleh pada Random Forest dengan SMOTE yang mencapai akurasi 86,96%, recall 87,25%, F1-score 88,12%, dan AUC-ROC 93,34%. Temuan ini menunjukkan bahwa kombinasi SMOTE dan optimasi hyperparameter mampu menghasilkan model klasifikasi penyakit jantung yang lebih seimbang, andal, dan potensial digunakan sebagai sistem pendukung keputusan klinis di layanan kesehatan.

Kata Kunci: Klasifikasi Penyakit Jantung; Data Imbalanced ; Random Forest; Xgboost; Lightgbm; Logistic Regression.

Abstract

This study examines the application of the Synthetic Minority Over-sampling Technique (SMOTE) for heart disease classification using four machine learning algorithms, namely Logistic Regression, Random Forest, LightGBM, and XGBoost, based on the Heart Disease UCI dataset consisting of 920 medical records with 16 clinical features. The original severity labels (0–4) are converted into two classes, namely not sick (0) and sick (1–4), to better align with binary decision-making needs in clinical screening. The experiments are conducted in two scenarios: (1) training models on the original data without handling class imbalance and (2) training models with SMOTE applied only to the training data within a pipeline, accompanied by hyperparameter tuning using k-fold cross-validation. Model performance is evaluated using accuracy, precision, recall, F1-score, AUC-ROC, as well as confusion matrix analysis to examine misclassifications, particularly false negatives in the sick class. In the scenario without SMOTE, the best model, Logistic Regression, achieves an accuracy of 84.78%, recall of 84.31%, F1-score of 86.00%, and AUC-ROC of 91.95%, although the number of false negatives remains relatively high. After applying SMOTE, there is an increase in recall and F1-score for the positive class across all models, with the best performance obtained by Random Forest with SMOTE, which achieves an accuracy of 86.96%, recall of 87.25%, F1-score of 88.12%, and AUC-ROC of 93.34%. These findings indicate that the combination of SMOTE and hyperparameter optimization can produce a more balanced and reliable heart disease classification model that is potentially useful as a clinical decision support system in healthcare services.

Keywords: Heart Disease Classification; Imbalanced Data; Random Forest; XGBoost; LightGBM; Logistic Regression.

1. PENDAHULUAN

Banyak jenis penyakit jantung, termasuk penyakit kardiovaskuler, jantung koroner, dan serangan jantung, didefinisikan sebagai penyakit jantung [1]. Penyakit jantung adalah salah satu penyakit kematian paling umum di dunia dengan berbagai gangguan jantung, beberapa kebiasaan yang tidak sehat, seperti kurangnya aktivitas fisik, diet yang tinggi lemak dan garam, merokok, konsumsi alkohol yang berlebihan, dan stres, adalah penyebab utama penyakit jantung [2]. Faktor-faktor tersebut sering kali muncul bersamaan dengan kondisi lain seperti obesitas, hipertensi, dan diabetes mellitus, sehingga semakin meningkatkan kemungkinan seseorang mengalami gangguan kardiovaskuler. Laporan World Health Organization (WHO) menunjukkan bahwa penyakit jantung (CVD) menyebabkan 17,7 juta kematian setiap tahun dan 31% dari semua kematian di dunia. Angka kematian ini diperkirakan akan terus meningkat dan diperkirakan akan mencapai 23,3 juta kematian pada tahun 2030, kejadian penyakit jantung dan pembuluh darah meningkat setiap tahun, menurut data Riset Kesehatan Dasar, dari 1000 orang, atau sekitar 2.784.064 orang di Indonesia, setidaknya 15 menderita penyakit jantung [3]. Ateroskelosis, proses kompleks pengendapan dan proliferasi komponen pada dinding arteri serta lipoprotein plasma, menyebabkan kerusakan sel otot jantung yang memompa dan membawa aliran darah ke seluruh tubuh serta Lipoprotein plasma berkembang dari kerak lemak sel busa menjadi timbunan plak yang ditutupi jaringan ikat, yang menghentikan aliran darah dan akhirnya menyebabkan peristiwa klinis. Pasien berusia sekitar 45 hingga 75 tahun paling sering menderita penyakit mematikan ini [2].

Kesehatan merupakan aspek yang paling penting bagi manusia karena setiap orang berpotensi mengalami masalah kesehatan, termasuk gangguan pada jantung. Salah satu bentuk gangguan tersebut adalah infeksi pada lapisan dalam jantung yang disebut endokarditis, yang umumnya disebabkan oleh infeksi virus atau bakteri. Salah satu bakteri yang

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9485
Hal 398-409
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

paling sering menyebabkan infeksi ini adalah *Streptococcus beta hemolyticus* [4]. Penanganan penyakit jantung dapat dilakukan melalui berbagai metode, termasuk tindakan operasi, prosedur intervensi, penyinaran, hingga kemoterapi pada kondisi tertentu. Namun, salah satu masalah utama dalam penanganan penyakit jantung adalah sulitnya deteksi dini. Banyak pasien baru menyadari kondisi mereka setelah muncul gejala yang berat, sehingga penyakit sudah berada pada tahap lanjut dan membutuhkan penanganan yang lebih kompleks [5]. Oleh karena itu, diperlukan suatu metode yang mampu membantu proses deteksi dan klasifikasi tingkat keparahan penyakit jantung sejak dini dan seakurat mungkin.

Diagnosis penyakit jantung umumnya melibatkan kombinasi informasi klinis, pemeriksaan fisik, hasil laboratorium, rekam jantung (EKG), serta pemeriksaan penunjang lain [6]. Namun, proses pengambilan keputusan klinis sering kali kompleks karena banyaknya variabel yang harus dipertimbangkan secara bersamaan, serta adanya variasi karakteristik pasien antar-populasi. Dalam konteks ini, pendekatan berbasis data dan *machine learning* menjadi menarik karena mampu memodelkan hubungan nonlinier dan kompleks antarkomponen klinis, serta memberikan prediksi risiko atau klasifikasi kondisi pasien secara lebih sistematis [7]. Untuk menjawab tantangan tersebut sekaligus memperkuat upaya deteksi dini dan pencegahan, penelitian ini memanfaatkan versi gabungan dataset UCI *Heart Disease* yang mengombinasikan data dari beberapa pusat (Cleveland, Hungarian, Switzerland, dan VA), sehingga diperoleh sekitar 920 data pasien dengan 16 fitur klinis utama. Setelah dilakukan proses prapemrosesan dan pengodean (*encoding*) variabel kategorikal, jumlah fitur yang digunakan model meningkat menjadi sekitar 29 fitur yang merepresentasikan karakteristik klinis pasien secara lebih rinci. Pendekatan ini diharapkan dapat memberikan gambaran yang lebih beragam mengenai profil pasien penyakit jantung, sehingga model yang dihasilkan lebih *robust* dan *generalizable* [8].

Fitur target pada dataset UCI *Heart Disease* pada awalnya berbentuk multikelas dengan nilai 0 hingga 4, yang menggambarkan tingkat keparahan penyakit jantung. Dalam praktik klinis, sering kali dibutuhkan keputusan biner yang lebih sederhana, yaitu membedakan antara tidak sakit (nilai 0) dan memiliki penyakit jantung (nilai 1-4). Oleh karena itu, pada penelitian ini label multikelas tersebut dikonversi menjadi label biner: kelas 0 sebagai "tidak sakit" dan kelas 1 sebagai "sakit" (gabungan label 1-4). Pendekatan ini sejalan dengan beberapa penelitian sebelumnya yang juga melakukan simplifikasi label untuk mendukung skenario skrining atau deteksi dini [7], [9].

Salah satu tantangan penting dalam klasifikasi penyakit jantung adalah ketidakseimbangan kelas (class imbalance). Pada dataset yang digunakan, proporsi pasien yang tidak sakit dan sakit tidak seimbang, sehingga model cenderung lebih "menghafal" kelas mayoritas dan mengabaikan kelas minoritas. Kondisi ini dapat menyebabkan nilai akurasi tampak tinggi, tetapi performa pada kelas pasien sakit (yang secara klinis justru lebih kritis) menjadi kurang optimal [9]. Untuk mengatasi masalah ini, teknik oversampling seperti *Synthetic Minority Over-sampling Technique* (SMOTE) sering digunakan. SMOTE bekerja dengan membangkitkan sampel sintetis baru di sekitar data minoritas yang ada, sehingga distribusi kelas menjadi lebih seimbang dan algoritme pembelajaran dapat mempelajari pola pada kelas minoritas dengan lebih baik [10].

Pada penelitian sebelumnya, Derisma dkk [11] membangun sistem prediksi penyakit jantung menggunakan dataset *Heart Disease* dari UCI *Repository* yang berisi 14 fitur numerik dan nominal dengan dua kelas, yaitu tidak adanya dan adanya penyakit jantung. Tiga algoritma klasifikasi yang dibandingkan adalah Naive Bayes, Random Forest, dan Neural Network, di mana seluruh proses mulai dari praproses, pelatihan hingga evaluasi dilakukan menggunakan *Orange* [12]. Kinerja model dievaluasi menggunakan metrik AUC, *classification accuracy* (CA), *F1-score*, *precision, recall*, *confusion matrix*, serta analisis kurva ROC. Hasil eksperimen menunjukkan bahwa Naive Bayes memberikan akurasi tertinggi sekitar 83%, sedikit lebih baik dibandingkan Random Forest (82%) dan Neural Network (81%), sehingga disimpulkan sebagai algoritma paling tepat untuk kasus prediksi penyakit jantung pada konfigurasi tersebut. Namun demikian, penelitian tersebut masih terbatas pada skenario klasifikasi biner tanpa eksplorasi penanganan ketidakseimbangan kelas, belum memanfaatkan teknik oversampling seperti SMOTE, dan belum mengkaji variasi algoritma ensemble modern maupun analisis performa sebelum dan sesudah balancing data sehingga ruang peningkatan akurasi dan kemampuan deteksi kasus positif masih terbuka.

Berdasarkan *research gap* tersebut, penelitian ini bertujuan mengembangkan model klasifikasi penyakit jantung yang lebih komprehensif dengan menggunakan empat algoritma pembelajaran mesin, yaitu Logistic Regression, Random Forest, LightGBM, dan XGBoost, pada dataset *Heart Disease* UCI [13]. Penelitian dirancang dalam dua skenario eksperimen: (1) pelatihan model tanpa SMOTE dan (2) pelatihan model dengan penerapan SMOTE pada data latih untuk mengatasi ketidakseimbangan kelas. Pada masing-masing skenario dilakukan hyperparameter tuning berbasis validasi silang sehingga setiap algoritma bekerja pada konfigurasi optimalnya. Kinerja model kemudian dievaluasi secara menyeluruh menggunakan metrik *accuracy*, *precision, recall*, *F1-score*, AUC-ROC, serta visualisasi confusion matrix untuk melihat kemampuan model dalam membedakan pasien sehat dan pasien berisiko. Dengan demikian, penelitian ini tidak hanya membandingkan beberapa algoritma, tetapi juga menganalisis dampak kombinasi SMOTE dan penyetelan hyperparameter terhadap peningkatan performa deteksi penyakit jantung, serta menentukan model terbaik yang berpotensi digunakan sebagai dasar sistem pendukung keputusan klinis.

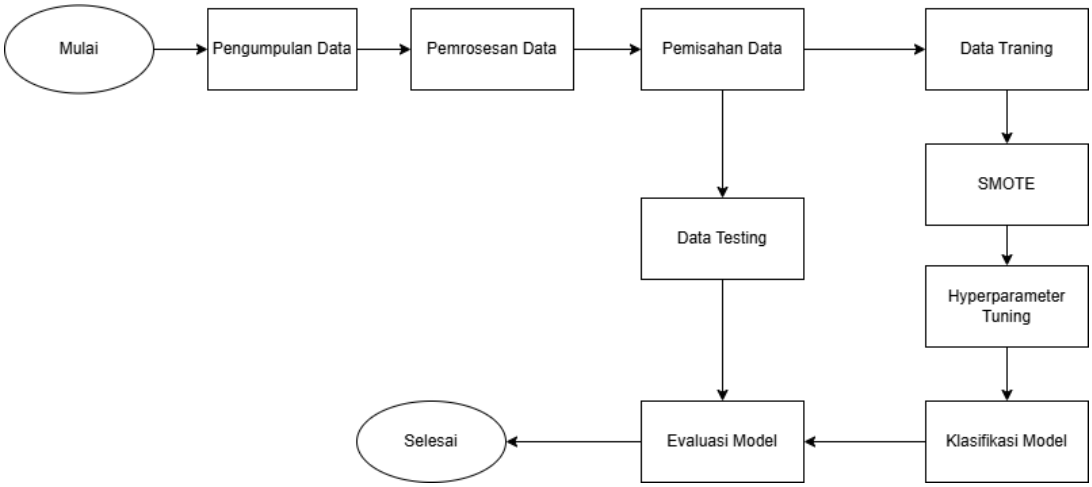
2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Gambar 1 menunjukkan tahapan penelitian yang digunakan dalam proses klasifikasi penyakit jantung. Penelitian dimulai dari tahap pengumpulan data, yaitu mengambil dataset *Heart Disease* UCI yang kemudian masuk ke tahap pemrosesan

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9485
Hal 398-409
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

data (pembersihan nilai hilang, pengubahan tipe data kategorikal menjadi numerik, serta normalisasi/standarisasi fitur). Setelah itu dilakukan pemisahan data menjadi data training dan data testing. Data *training* selanjutnya diproses dengan SMOTE untuk menyeimbangkan distribusi kelas, lalu dilakukan *hyperparameter tuning* pada keempat algoritma (Logistic Regression, Random Forest, LightGBM, dan XGBoost) hingga diperoleh konfigurasi model terbaik. Model yang sudah terlatih kemudian digunakan pada tahap klasifikasi model untuk memprediksi label pada data testing. Hasil prediksi tersebut dievaluasi pada tahap evaluasi model menggunakan metrik akurasi, presisi, recall, F1-score, dan AUC-ROC, sehingga pada akhirnya dapat ditarik kesimpulan mengenai kinerja masing-masing model dan pengaruh penerapan SMOTE dalam penelitian ini.



Gambar 1. Tahapan Penelitian

2.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari *Heart Disease Dataset* pada repositori UCI [14]. Dataset ini berisi data rekam medis pasien yang menjalani pemeriksaan terkait penyakit jantung di beberapa rumah sakit, yaitu Cleveland, Hungarian, Switzerland, dan VA Long Beach. Informasi asal rumah sakit disimpan dalam atribut dataset, sehingga setiap baris data dapat ditelusuri sumbernya. Secara keseluruhan, dataset terdiri dari 920 baris data dan 16 fitur yang mencakup 15 fitur prediktor dan 1 fitur target. Fitur prediktor yang digunakan antara lain*:* *age* (usia pasien), *sex* (jenis kelamin), *cp* (tipe nyeri dada), *trestbps* (tekanan darah saat istirahat), *chol* (kadar kolesterol total), *fbs* (kadar gula darah puasa), *restecg* (hasil elektrokardiogram saat istirahat), *thalch* (detak jantung maksimum saat latihan), *exang* (angina yang diinduksi oleh latihan), *oldpeak* (depresi ST), *slope* (kemiringan segmen ST), *ca* (jumlah pembuluh darah utama yang terdeteksi dengan fluoroskopi), thal (status thalassemia), dataset (kode sumber rumah sakit), serta num sebagai atribut target yang menunjukkan tingkat keparahan penyakit jantung.

2.3 *Data pre-processing*

Selanjutnya dilakukan analisis kualitas data untuk mengidentifikasi keberadaan nilai hilang (*missing value*) dan inkonsistensi tipe data. Beberapa fitur, khususnya *ca* dan *thal*, mengandung nilai hilang yang terlebih dahulu dikonversi menjadi *NaN* sehingga dapat diproses sebagai nilai hilang oleh pustaka *scikit-learn* [15] . Jumlah nilai hilang pada setiap kolom dihitung, dan karena proporsinya masih dalam batas wajar, seluruh baris tetap dipertahankan dengan menerapkan teknik imputasi. Untuk atribut numerik seperti *age, trestbps, chol, thalach,* dan *oldpeak*, digunakan imputasi median agar tidak terlalu dipengaruhi oleh nilai pencilan, sejalan dengan rekomendasi pada penelitian klasifikasi penyakit jantung dan dataset klinis yang cenderung mengandung outlier [9], [10].Sementara itu, atribut kategorikal seperti *cp, restecg, slope, ca,* dan *thal* diimputasi menggunakan modus (*most frequent*) sehingga kategori yang paling sering muncul menjadi nilai pengganti [10]. Pendekatan kombinasi imputasi median dan modus ini banyak digunakan pada riset-riset kesehatan karena sederhana, stabil, dan tidak mengubah distribusi data secara drastis [9], [10], [16].

Setelah penanganan nilai hilang, dilakukan transformasi fitur kategorikal menjadi representasi numerik. Beberapa atribut pada dataset, seperti *sex, cp, fbs, restecg, exang, slope, ca,* dan *thal*, secara konseptual bersifat kategorikal meskipun pada file asli disajikan dalam bentuk kode bilangan bulat. Untuk menghindari makna urutan yang tidak semestinya dan mendukung algoritma pembelajaran mesin, fitur-fitur ini dikodekan menggunakan *One-Hot Encoding* sehingga setiap kategori direpresentasikan sebagai vektor biner [17], [18], [19].

2.4 Pemisahan Data

Tabel 1 menunjukan proporsi pembagian dataset. Pada tahap ini, dataset dibagi menjadi dua bagian utama, yaitu data latih (*training set*) dan data uji (*testing set*). Data latih digunakan untuk melatih model agar mampu mempelajari pola hubungan antara fitur-fitur klinis dengan label penyakit jantung, sedangkan data uji digunakan untuk mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9485
Hal 398-409
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Tabel 1. Pemisahan Data

Jenis Data	Jumlah Data	Presentase (%)
Data Latih	736	80
Data Uji	184	20
Total	920	100

Dengan total 920 data, pembagian ini menghasilkan sekitar 736 data latih dan 184 data uji. Proses pembagian dilakukan secara acak menggunakan parameter *random_state* tertentu agar hasilnya dapat direproduksi. Selain itu, pembagian data dilakukan secara stratified, yaitu mempertahankan proporsi kelas pada data latih dan data uji agar distribusi kelas tidak berubah signifikan.

2.5 Penanganan Data Imbalanced Menggunakan SMOTE

Synthetic Minority Over-sampling Technique (SMOTE) digunakan dalam penelitian ini untuk mengatasi masalah ketidakseimbangan kelas yang umum terjadi pada data medis, termasuk dataset penyakit jantung, di mana jumlah pasien "tidak sakit" jauh lebih banyak dibandingkan pasien "sakit". Ketidakseimbangan ini dapat membuat model cenderung bias ke kelas mayoritas sehingga sulit mengenali pasien berisiko. Berbeda dengan *random oversampling* yang hanya menduplikasi data, SMOTE membangkitkan sampel sintetis baru pada kelas minoritas dengan melakukan interpolasi antara suatu sampel dan beberapa tetangga terdekatnya di ruang fitur, sehingga distribusi kelas menjadi lebih seimbang tanpa terlalu meningkatkan risiko *overfitting* [20], [21], [22]. Dalam berbagai penelitian prediksi penyakit jantung, SMOTE terbukti efektif meningkatkan *recall*, F1-score, dan AUC ketika dikombinasikan dengan algoritma seperti Random Forest, XGBoost, dan berbagai metode *ensemble* lainnya, bahkan sering dikembangkan bersama teknik pembersihan data seperti SMOTE-ENN untuk menghasilkan batas keputusan yang lebih bersih [9], [10], [20], [23]. Pada Tabel 2, SMOTE diterapkan pada data latih setelah pra-pemrosesan dan sebelum pelatihan model untuk menyeimbangkan distribusi kelas multikelas (0-4), sehingga keempat algoritma; Logistic Regression, Random Forest, LightGBM, dan XGBoost diharapkan dapat mempelajari pola pasien "sakit" dengan lebih baik dan memberikan performa klasifikasi yang lebih seimbang antara kelas negatif dan positif.

Tabel 2. Penanganan *Data Imbalanced* menggunakan SMOTE

Jenis Data	Kelas	Tanpa SMOTE	Dengan SMOTE
Data Latih	0	329	329
	1	212	329
	2	87	329
	3	86	329
	4	22	329

2.6 *Hyperparameter Tuning*

Pada penelitian ini, proses *hyperparameter tuning* dilakukan untuk meningkatkan kinerja empat algoritma klasifikasi yang digunakan, yaitu Logistic Regression, Random Forest, XGBoost, dan LightGBM. *Hyperparameter* adalah parameter yang ditentukan sebelum proses pelatihan, seperti nilai regularisasi C pada Logistic Regression, jumlah pohon (*n_estimators*) dan kedalaman maksimum (*max_depth*) pada Random Forest, serta *learning_rate*, jumlah pohon, dan parameter regulasi lainnya pada XGBoost dan LightGBM. Pemilihan kombinasi hyperparameter yang tepat penting untuk mengurangi risiko *underfitting* maupun *overfitting* pada data klinis yang kompleks dan berpotensi mengandung pencilan [8], [18], [24], [25]. Strategi tuning dilakukan secara sistematis menggunakan skema pencarian berbasis validasi silang (*k-fold cross-validation*) pada data latih. Untuk setiap kombinasi hyperparameter, model dilatih dan divalidasi pada beberapa lipatan, kemudian rata-rata metrik performa digunakan untuk menilai kualitas kombinasi tersebut. Pendekatan ini sejalan dengan berbagai studi terkini yang menunjukkan bahwa optimasi hyperparameter melalui prosedur terstruktur, seperti *randomize cv* atau metode optimasi lainnya, mampu meningkatkan akurasi dan stabilitas model pada tugas prediksi penyakit jantung dan gagal jantung [17], [26], [27], [28], [29]. Dalam penelitian ini, tuning dilakukan pada dua skenario, yaitu tanpa SMOTE dan dengan SMOTE, di mana pada skenario kedua, SMOTE diintegrasikan ke dalam *pipeline* bersama dengan model klasifikasi. Pendekatan ini mengikuti tren penelitian yang mengombinasikan optimasi hyperparameter dan penanganan ketidakseimbangan kelas untuk menghasilkan model yang lebih seimbang dalam mengenali pasien dengan dan tanpa penyakit jantung [9], [20], [30].

2.7 Klasifikasi Model

Pada tahap klasifikasi model, Logistic Regression digunakan sebagai model dasar (*baseline*) karena bersifat sederhana, interpretable, dan banyak dipakai pada data klinis tabular untuk mengestimasi probabilitas kejadian penyakit jantung. Model dilatih menggunakan fitur hasil pra-pemrosesan (imputasi, *one-hot encoding*, dan standardisasi) pada data latih, baik untuk skenario label multikelas maupun label biner. Kombinasi hyperparameter terbaik, seperti nilai regularisasi C, jenis penalti, dan *solver*, diambil dari hasil proses *hyperparameter tuning* sehingga model tidak terlalu overfitting namun tetap mampu menangkap pola penting pada data [8], [17], [18] . Logistic Regression dilatih pada dua skenario, yaitu tanpa

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9485
Hal 398-409
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

SMOTE dan dengan SMOTE pada data latih, sehingga dapat dibandingkan pengaruh penyeimbangan kelas terhadap stabilitas dan kemampuan generalisasi model [9], [10], [17].

Random Forest diterapkan sebagai model *ensemble* berbasis banyak pohon keputusan yang mampu memodelkan hubungan non-linier dan interaksi antar fitur pada data penyakit jantung. Pada penelitian ini, Random Forest dilatih menggunakan kombinasi hyperparameter hasil tuning, seperti jumlah pohon (*n_estimators*), kedalaman maksimum (*max_depth*), jumlah fitur di setiap pemisahan (*max_features*), serta batas minimum sampel pada simpul internal dan daun. Pengaturan ini mengontrol kompleksitas dan keragaman pohon sehingga diperoleh keseimbangan antara bias dan varians [29], [31], [32]. Sama seperti Logistic Regression, Random Forest dilatih pada skenario tanpa SMOTE dan dengan SMOTE yang hanya diterapkan pada data latih, mengikuti rekomendasi penelitian yang menunjukkan bahwa kombinasi Random Forest dan teknik oversampling seperti SMOTE atau SMOTE-ENN dapat meningkatkan deteksi kelas minoritas pada kasus penyakit jantung [9], [20], [22], [26].

LightGBM digunakan sebagai representasi algoritma *gradient boosting* generasi baru yang dirancang efisien pada data tabular dan mampu menangani fitur dengan distribusi kompleks. Model ini dilatih dengan memanfaatkan hyperparameter hasil tuning, seperti *learning_rate*, jumlah pohon, *num_leaves*, *max_depth*, serta parameter regularisasi seperti *subsample* dan *feature_fraction*. Penggunaan LightGBM pada prediksi penyakit jantung dan gagal jantung dalam berbagai penelitian menunjukkan bahwa kombinasi pengaturan *learning_rate* yang kecil, jumlah iterasi yang cukup, dan regularisasi yang tepat dapat menghasilkan model dengan akurasi tinggi namun tetap terkendali terhadap *overfitting* [17], [24], [30], [33]. Pada penelitian ini, LightGBM juga dilatih dalam dua skenario, yaitu tanpa SMOTE dan dengan SMOTE dalam sebuah *pipeline*, sehingga efek penyeimbangan kelas terhadap performa *gradient boosting* dapat dianalisis secara sistematis [20], [21], [23], [28], [33].

XGBoost dipilih sebagai algoritma *gradient boosting* lain yang telah banyak dilaporkan berkinerja sangat baik pada tugas prediksi penyakit jantung, terutama ketika dikombinasikan dengan pemilihan fitur dan teknik resampling seperti SMOTE [16], [30], [34]. Pada tahap pelatihan, XGBoost dikonfigurasi menggunakan hyperparameter terbaik hasil tuning, meliputi *learning_rate*, jumlah pohon, *max_depth*, *min_child_weight*, serta parameter regulasi seperti *subsample* dan *colsample_bytree* untuk mengendalikan kompleksitas model dan mencegah overfitting [16], [24], [25]. Sejalan dengan LightGBM, XGBoost dilatih pada data latih asli (tanpa SMOTE) dan pada data latih yang telah di-*oversampling* dengan SMOTE di dalam *pipeline*, dengan SMOTE hanya diterapkan pada data latih di setiap lipatan validasi silang agar distribusi data uji tetap merepresentasikan kondisi ketidakseimbangan kelas yang nyata [9], [10], [22].

2.8 Evaluasi Model

Evaluasi kinerja model dilakukan menggunakan data uji yang tidak pernah digunakan selama proses pelatihan maupun tuning, baik untuk label multikelas maupun label biner. Penilaian didasarkan pada beberapa metrik yang umum digunakan dalam penelitian prediksi penyakit jantung, yaitu akurasi, presisi, *recall*, F1-*score* berbobot (F1-*weighted*), serta Area Under the ROC Curve (AUC-ROC) untuk skenario biner [18], [25], [27]. Untuk kasus biner, metrik-metrik tersebut dihitung berdasarkan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) yang dirangkum dalam *confusion matrix*. Rumus masing-masing metrik didefinisikan sebagai berikut [10], [25], [34]:

a. *Accuracy* (Akurasi)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

b. *Precision* (Presisi)

$$Precision = \frac{TP}{TP+TP} \quad (2)$$

c. *Recall*

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

d. *F1-Score*

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Pada skenario multikelas, F1-*score* dihitung dalam bentuk F1-*score* berbobot (F1-*weighted*), yaitu rata-rata F1 per kelas yang ditimbang dengan proporsi jumlah sampel tiap kelas. Pendekatan ini dipilih karena lebih representatif ketika terdapat ketidakseimbangan jumlah data antar kelas dan banyak digunakan dalam penelitian yang mengevaluasi model klasifikasi penyakit jantung dan gagal jantung [24], [25]. Selain itu, untuk label biner digunakan kurva ROC dan nilai AUC-ROC untuk menilai kemampuan model membedakan kelas "sakit" dan "tidak sakit" pada berbagai ambang keputusan (*threshold*). Studi-studi sebelumnya menunjukkan bahwa kombinasi F1-*score*, akurasi, dan AUC-ROC memberikan gambaran yang lebih komprehensif terhadap kinerja model pada data klinis yang tidak seimbang [16], [18], [27]. Hasil evaluasi pada metrik-metrik tersebut kemudian digunakan untuk membandingkan kinerja antarmodel, serta antara skenario tanpa SMOTE dan dengan SMOTE.

e. *ROC* dan *AUC*

Sumbu Y*: True Positive Rate* (TPR) = $TPR \frac{TP}{TP + FN}$ (5)

Sumbu X: *False Positive Rate* (FPR) = $FPR \frac{FP}{FP + TN}$ (6)

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9485
Hal 398-409
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Nilai AUC = $AUC \int_0^1 TPR(FPR)d(FPR)$ (7)

Keterangan:
TP (*True Positive*): Jumlah data positif (TB) yang diprediksi benar sebagai positif.
TN (*True Negative*): Jumlah data negatif (Normal) yang diprediksi benar sebagai negatif.
FP (*False Positive*): Jumlah data negatif yang salah diprediksi sebagai positif.
FN (*False Negative*): Jumlah data positif yang salah diprediksi sebagai negatif.

3. HASIL DAN PEMBAHASAN

3.1 Tahapan Penerapan Algoritma

Tahapan penerapan algoritma dalam penelitian ini dilakukan secara sistematis untuk memastikan proses eksperimen berjalan terstruktur dan dapat direplikasi. Proses dimulai dari tahap pra-pemrosesan data yang meliputi penanganan nilai hilang (imputasi), *encoding* variabel kategorikal, serta standardisasi fitur numerik guna menyamakan skala antarvariabel. Tahap ini bertujuan untuk meningkatkan stabilitas dan performa model pada saat pelatihan. Selanjutnya, dataset dibagi menjadi data latih dan data uji menggunakan rasio pembagian tertentu. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

Penelitian ini menerapkan dua skenario eksperimen. Pada skenario pertama, model dilatih tanpa penanganan ketidakseimbangan kelas untuk melihat performa awal pada distribusi data asli. Pada skenario kedua, diterapkan metode *Synthetic Minority Oversampling Technique* (SMOTE) pada data latih guna mengatasi ketidakseimbangan kelas sebelum proses pelatihan model dilakukan. Empat algoritma klasifikasi yang digunakan dalam penelitian ini adalah Logistic Regression, Random Forest, LightGBM, dan XGBoost. Masing-masing model dilatih menggunakan konfigurasi parameter yang telah disesuaikan untuk memperoleh performa optimal. Evaluasi kinerja dilakukan menggunakan metrik akurasi, precision, recall, F1-*score*, serta AUC-ROC. Selain itu, analisis confusion matrix juga digunakan untuk mengidentifikasi pola kesalahan klasifikasi, khususnya pada kasus false negative yang kritis dalam konteks deteksi penyakit jantung.

3.2 Evaluasi Model Tanpa SMOTE

Ringkasan hasil evaluasi tersebut ditunjukkan pada Tabel 3. Pada hasil evaluasi kinerja keempat model klasifikasi Logistic Regression, LightGBM, XGBoost, dan Random Forest pada skenario tanpa penerapan SMOTE. Seluruh model dilatih menggunakan label multikelas tingkat keparahan penyakit jantung (0-4), kemudian dievaluasi pada data testing dalam bentuk label biner, yaitu kelas 0 sebagai "tidak sakit" dan kelas 1-4 sebagai "sakit", dengan tingkat ringan, moderat, serius, dan sangat serius. Kinerja masing-masing model diukur menggunakan metrik akurasi, precision, recall, F1-Score, dan AUC-ROC untuk melihat kemampuan model dalam membedakan pasien sehat dan pasien dengan level penyakit jantung.

Tabel 3. Hasil Evaluasi Tanpa SMOTE di *Data testing*

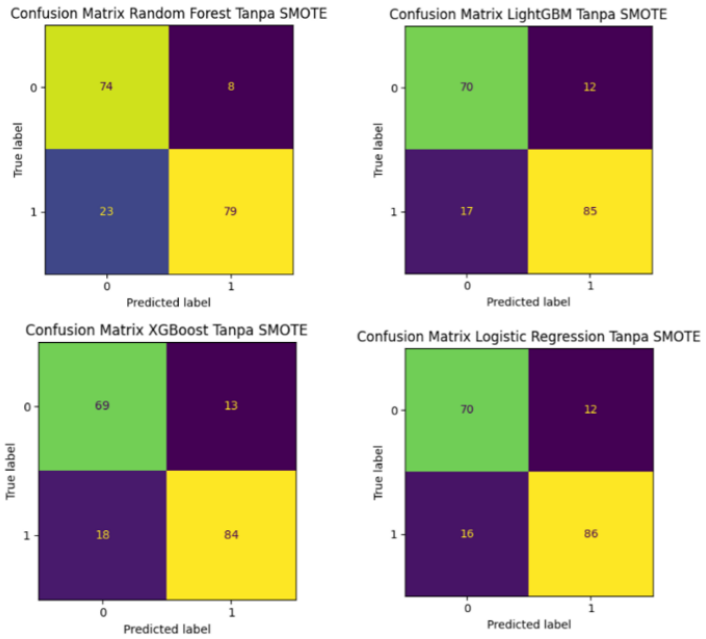
Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	0,8478	0,8775	0,8431	0,8600	0,9195
LightGBM	0,8423	0,8762	0,8333	0,8542	0,9091
XGBoost	0,8315	0,8659	0,8235	0,8442	0,9074
Random Forest	0,8315	0,9080	0,7745	0,8359	0,9310

Tabel 3 menyajikan ringkasan hasil evaluasi empat algoritma klasifikasi Logistic Regression, LightGBM, XGBoost, dan Random Forest pada skenario tanpa penerapan SMOTE. Seluruh model dilatih menggunakan label multikelas (0-4), kemudian dievaluasi dalam bentuk label biner (0 = tidak sakit, 1 = sakit). Jika dibandingkan dengan penelitian Derisma yang menggunakan dataset *Heart Disease* UCI Cleveland dan menemukan bahwa Naive Bayes merupakan model terbaik dengan akurasi sekitar 83% berdasarkan evaluasi accuracy, precision, recall, F1-score, dan AUC-ROC, hasil eksperimen tanpa SMOTE pada penelitian ini menunjukkan kinerja yang umumnya lebih tinggi dan lebih stabil di keempat algoritma yang diuji. Logistic Regression mencapai akurasi 0,8478 dengan precision 0,8775, recall 0,8431, F1-score 0,8600, dan AUC-ROC 0,9195; LightGBM menghasilkan akurasi 0,8423, precision 0,8762, recall 0,8333, F1-score 0,8542, dan AUC-ROC 0,9091; XGBoost memperoleh akurasi 0,8315, precision 0,8659, recall 0,8235, F1-score 0,8442, dan AUC-ROC 0,9074; sedangkan Random Forest mencatat akurasi 0,8315 dengan precision tertinggi 0,9080, recall 0,7745, F1-score 0,8359, dan AUC-ROC terbaik 0,9310. Secara keseluruhan, hampir semua model pada penelitian ini mampu menyamai atau melampaui akurasi terbaik yang dilaporkan Derisma, sekaligus menunjukkan kombinasi precision yang lebih tinggi dan AUC-ROC di atas 0,90 yang mengindikasikan kemampuan pemisahan kelas sehat dan sakit yang lebih kuat; perbedaan utama terletak pada trade-off antara recall dan precision, di mana model-model pada penelitian ini cenderung mengutamakan pengurangan false positive sekaligus mempertahankan tingkat deteksi pasien sakit yang kompetitif dibandingkan studi Derisma dkk.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9485
Hal 398-409
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

3.2 Analisis *Confusion Matrix* Tanpa SMOTE

Analisis pola kesalahan klasifikasi masing-masing model melalui confusion matrix pada skenario tanpa SMOTE. Confusion matrix menggambarkan sebaran prediksi benar dan salah untuk dua kelas, yaitu "tidak sakit" (0) dan "sakit" (1), sehingga memudahkan identifikasi nilai true positive, true negative, false positive, dan terutama false negative yang sangat kritis dalam konteks diagnosis penyakit jantung. Gambar 2 menampilkan confusion matrix untuk keempat model Random Forest, LightGBM, XGBoost, dan Logistic Regression sebagai dasar untuk menilai sejauh mana setiap algoritma mampu mendeteksi pasien yang benar-benar sakit tanpa mengorbankan terlalu banyak kesalahan pada kelas tidak sakit.



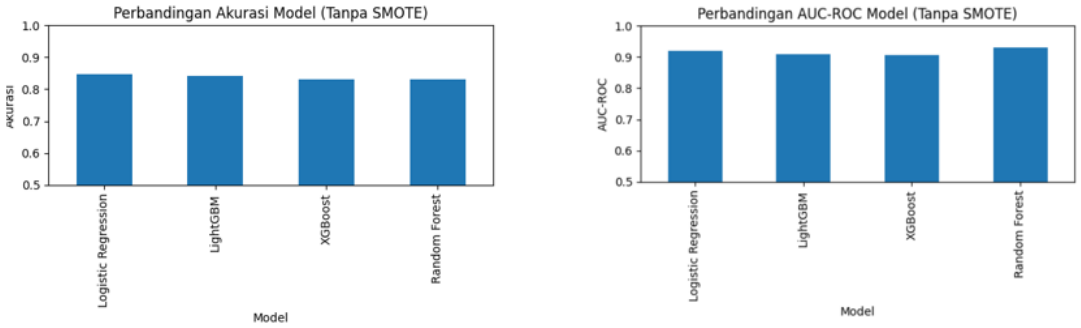
Gambar 2. *Confusion matrix* untuk Keempat Model: LR, RF,LightGBM dan XGBoost

Gambar 2 memperlihatkan *confusion matrix* masing-masing model pada skenario tanpa SMOTE. *Confusion matrix* Logistic Regression menunjukkan 70 sampel negatif yang diklasifikasikan benar (*true negative*), 12 sampel negatif yang salah diprediksi sebagai positif (*false positive*), 86 sampel positif yang diklasifikasikan benar (*true positive*), dan 16 sampel positif yang salah diklasifikasikan sebagai negatif (*false negative*); artinya, masih terdapat sejumlah pasien yang sebenarnya sakit tetapi tidak terdeteksi oleh model. LightGBM menampilkan pola yang sangat mirip, dengan 70 *true negative* dan 12 *false positive*, namun *true positive* sedikit menurun menjadi 85 dan *false negative* meningkat menjadi 17, sejalan dengan nilai *recall* LightGBM yang sedikit lebih rendah dibandingkan Logistic Regression sehingga lebih sering melewatkan pasien sakit. Pada XGBoost, jumlah *true negative* turun menjadi 69, *false positive* naik menjadi 13, *true positive* menjadi 84, dan *false negative* meningkat menjadi 18; hal ini menjelaskan mengapa *recall* dan F1-score XGBoost sedikit lebih rendah dibanding dua model sebelumnya. Random Forest menghasilkan *true negative* terbanyak (74) dengan *false positive* paling sedikit (8), tetapi memiliki *false negative* tertinggi (23), sehingga tanpa SMOTE model ini cenderung konservatif: sangat jarang menandai pasien sehat sebagai sakit, namun konsekuensinya lebih banyak pasien sakit yang tidak terdeteksi. Jika dibandingkan secara umum dengan penelitian Derisma dkk. di mana model terbaik Naive Bayes pada 303 data uji menghasilkan 139 *true negative*, 112 *true positive*, 25 *false positive*, dan 27 *false negative* dengan akurasi sekitar 83% struktur *confusion matrix* Logistic Regression pada penelitian ini relatif lebih seimbang, karena pada 184 data uji hanya terdapat 12 *false positive* dan 16 *false negative* (sekitar 16 dari 102 pasien sakit), sehingga secara proporsional sensitivitas terhadap pasien sakit sedikit lebih baik meskipun tingkat kesalahan *false positive* hampir serupa. Sementara itu, pola Random Forest pada kedua penelitian sama-sama menunjukkan kecenderungan menekan *false positive* dengan konsekuensi meningkatnya *false negative*. Secara keseluruhan, baik hasil penelitian ini maupun Derisma dkk. sama-sama mengindikasikan bahwa masalah *false negative* masih menjadi tantangan utama dalam klasifikasi penyakit jantung, sehingga diperlukan teknik penanganan ketidakseimbangan kelas seperti SMOTE untuk meningkatkan kemampuan model dalam mendeteksi pasien yang benar-benar sakit.

3.3 Perbandingan Akurasi dan AUC-ROC Tanpa SMOTE

Perbandingan kinerja keempat model pada skenario tanpa SMOTE berdasarkan dua metrik utama, yaitu akurasi dan AUC-ROC. Akurasi menggambarkan proporsi prediksi yang benar terhadap seluruh data uji, sedangkan AUC-ROC menunjukkan kemampuan model membedakan antara pasien sakit dan tidak sakit pada berbagai ambang keputusan. Gambar 3 menampilkan grafik batang akurasi dan AUC-ROC untuk Logistic Regression, LightGBM, XGBoost, dan Random Forest, sehingga memudahkan identifikasi model mana yang paling konsisten memberikan prediksi benar sekaligus memiliki kemampuan diskriminasi terbaik tanpa penerapan teknik penyeimbangan kelas.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9485
Hal 398-409
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 3. Perbandingan Akurasi dan AUC-ROC Tanpa SMOTE

Perbandingan akurasi dan AUC-ROC antar model pada skenario tanpa SMOTE divisualisasikan pada Gambar 3. Pada grafik perbandingan akurasi model (tanpa SMOTE), tampak bahwa semua model memiliki akurasi yang sangat mirip di kisaran 0,83-0,85; Logistic Regression sedikit lebih unggul dengan akurasi 0,84, diikuti LightGBM, XGBoost, dan Random Forest dengan selisih yang relatif kecil, sehingga jika hanya dilihat dari akurasi, perbedaan performa keempat model belum terlalu menonjol. Sementara pada grafik perbandingan AUC-ROC (tanpa SMOTE)" memberikan perspektif lain: Random Forest menempati posisi teratas dengan AUC 0,93, disusul Logistic Regression 0,92 serta LightGBM dan XGBoost di 0,91, yang menunjukkan bahwa seluruh model memiliki kemampuan diskriminatif yang sangat baik dalam membedakan pasien sehat dan sakit pada berbagai ambang keputusan. Jika dibandingkan dengan penelitian Derisma dkk., di mana model terbaik Naive Bayes hanya mencapai akurasi sekitar 83% tanpa pelaporan AUC-ROC yang rinci, maka keempat model pada penelitian ini khususnya Logistic Regression dan Random Forest menunjukkan peningkatan akurasi sekaligus kemampuan pemisahan kelas yang lebih kuat. Dengan demikian, pada skenario tanpa SMOTE dapat disimpulkan bahwa Logistic Regression unggul dari sisi akurasi dan F1-score, sementara Random Forest memiliki kelebihan pada precision dan AUC-ROC; namun, sebagaimana juga terlihat pada penelitian Derisma dkk., kedua penelitian sama-sama masih menghadapi keterbatasan dalam menekan kasus *false negative*, sehingga penyeimbangan data melalui teknik seperti SMOTE menjadi langkah penting pada eksperimen berikutnya.

3.4 Evaluasi Model Dengan SMOTE

Hasil evaluasi keempat model klasifikasi setelah dilakukan penyeimbangan kelas menggunakan metode SMOTE pada data latih. Nilai akurasi, precision, recall, F1-score, dan AUC-ROC yang ditampilkan pada Tabel 4 dihitung berdasarkan prediksi model terhadap data *testing* yang tidak mengalami *oversampling*, sehingga dapat menggambarkan kemampuan generalisasi setiap algoritma dalam mendeteksi pasien penyakit jantung pada kondisi distribusi data yang sebenarnya.

Tabel 4. Evaluasi Model dengan SMOTE *Data Testing*

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	0,8369	0,8396	0,8725	0,8557	0,9032
LightGBM	0,8532	0,8504	0,8921	0,8708	0,9338
XGBoost	0,8423	0,8411	0,8823	0,8612	0,9310
Random Forest	0,8695	0,8900	0,8725	0,8811	0,9334

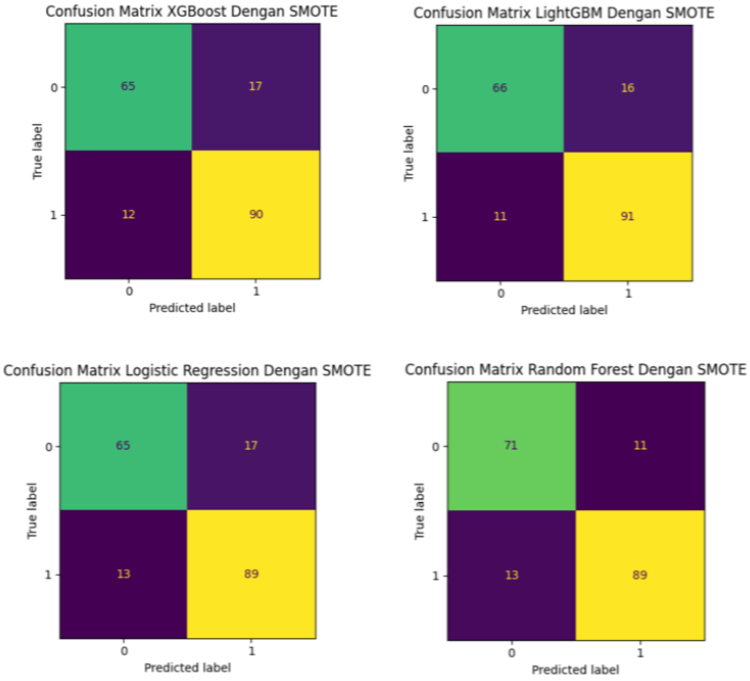
Tabel 4 menyajikan hasil evaluasi keempat model setelah penerapan SMOTE pada data latih. Berbeda dengan skenario sebelumnya, distribusi kelas pada data latih telah dibuat seimbang, sementara data uji tetap mempertahankan distribusi aslinya. Secara umum, penerapan SMOTE menghasilkan peningkatan recall dan F1-score, khususnya pada model-model *ensemble tree*. Random Forest menjadi model dengan performa terbaik secara keseluruhan, dengan akurasi sebesar 0,8696, precision sekitar 0,8900, recall 0,8725, F1-score 0,8812, dan AUC-ROC 0,9334. Angka-angka ini lebih tinggi dibandingkan hasil Random Forest tanpa SMOTE, terutama pada metrik recall dan F1-score. LightGBM juga menunjukkan peningkatan yang jelas, dengan akurasi 0,8533, recall 0,8922, F1-score 0,8708, dan AUC-ROC 0,9339. XGBoost mengalami kenaikan akurasi dan F1-score menjadi sekitar 0,8424 dan 0,8612, dengan AUC-ROC 0,9310, sehingga tetap kompetitif. Logistic Regression sedikit menurun pada sisi akurasi (dari 0,8478 menjadi 0,8370), namun recall meningkat dari 0,8431 menjadi 0,8725, sehingga F1-score tetap berada di kisaran 0,8557. Artinya, model menjadi lebih sensitif terhadap kelas positif meskipun mengorbankan sedikit akurasi. Secara keseluruhan, perbandingan Tabel 3 dan Tabel 4 menunjukkan bahwa SMOTE membantu model lebih baik dalam mendeteksi pasien yang benar-benar sakit, terutama untuk LightGBM, XGBoost, dan Random Forest.

3.5 Analisis Confusion Matrix Dengan SMOTE

Analisis lebih lanjut terhadap pola kesalahan klasifikasi model setelah penerapan SMOTE pada data latih. Gambar 4 menampilkan *confusion matrix* untuk keempat algoritma, yaitu Logistic Regression, Random Forest, LightGBM, dan XGBoost, sehingga dapat diamati perbandingan jumlah *true positive*, *true negative*, *false positive*, dan *false negative* pada

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9485
Hal 398-409
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

masing-masing model. Fokus utama analisis ditujukan pada penurunan kasus *false negative* (pasien sakit yang diprediksi tidak sakit) setelah oversampling, karena kesalahan jenis ini paling berisiko dalam konteks deteksi dini penyakit jantung.

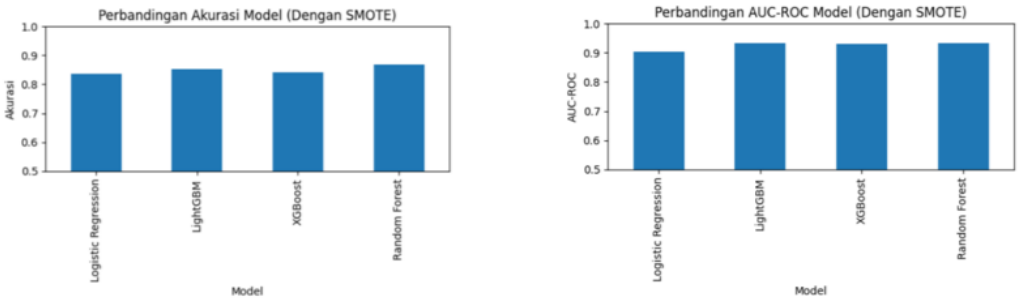


Gambar 4. *Confusion matrix* dengan SMOTE untuk Keempat Model.

Gambar 4 memperlihatkan perubahan pola *confusion matrix* keempat model setelah penerapan SMOTE jika dibandingkan dengan kondisi tanpa SMOTE. Pada Logistic Regression, kombinasi nilai 65 TN, 17 FP, 89 TP, dan 13 FN menunjukkan bahwa jumlah pasien sakit yang berhasil terdeteksi meningkat (TP naik dari 86 menjadi 89 dan FN turun dari 16 menjadi 13), sehingga recall/sensitivitas naik dari sekitar 0,84 menjadi 0,87 meskipun akurasi sedikit menurun karena bertambahnya FP. LightGBM mengalami perbaikan yang lebih jelas: dari (70 TN, 12 FP, 85 TP, 17 FN) menjadi (66 TN, 16 FP, 91 TP, 11 FN), yang berarti model jauh lebih sensitif terhadap pasien sakit recall naik dari 0,83 menjadi 0,89 dengan kompromi berupa sedikit peningkatan kesalahan prediksi pada pasien sehat. Pola yang sama terlihat pada XGBoost dan Random Forest: keduanya mengalami penurunan FN (XGBoost dari 18 menjadi 12, Random Forest dari 23 menjadi 13) sehingga recall meningkat tajam RF dari 0,77 ke 0,87, sementara kenaikan FP masih relatif kecil. Jika dibandingkan dengan hasil tanpa SMOTE, jelas bahwa perbaikan utama bukan sekadar pada akurasi dan AUC-ROC, melainkan pada nilai recall yang mencerminkan sensitivitas model dalam menangkap kasus positif; dalam konteks medis, peningkatan sensitivitas ini sangat penting karena lebih baik sedikit "berlebihan" menandai pasien sebagai berisiko (FP naik moderat) daripada melewatkan pasien yang benar-benar sakit (FN tinggi) seperti yang terjadi pada skenario tanpa SMOTE.

3.6 Perbandingan Akurasi dan AUC-ROC Dengan SMOTE

Perbandingan nilai akurasi dan AUC-ROC dari keempat algoritma klasifikasi. Gambar 5 memperlihatkan bagaimana perubahan distribusi data latih melalui oversampling memengaruhi kemampuan model dalam mengklasifikasikan pasien sakit dan tidak sakit secara keseluruhan (akurasi), sekaligus membedakan kedua kelas pada berbagai ambang keputusan (AUC-ROC). Melalui visualisasi ini, dapat terlihat model mana yang memperoleh peningkatan kinerja paling konsisten setelah SMOTE, serta seberapa besar perbedaan performanya dibandingkan algoritma lain pada skenario data seimbang.



Gambar 5. Perbandingan Akurasi dan AUC-ROC Dengan SMOTE

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9485
Hal 398-409
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Perbandingan akurasi dan AUC-ROC model setelah penerapan SMOTE ditunjukan pada Gambar 5. Pada grafik perbandingan akurasi model (dengan SMOTE)" terlihat bahwa Random Forest memiliki akurasi tertinggi sekitar 0,87, diikuti LightGBM sekitar 0,85, sementara XGBoost dan Logistic Regression berada di kisaran 0,84; kenaikan batang Random Forest dan LightGBM dibandingkan grafik tanpa SMOTE menunjukkan bahwa kedua model ini paling diuntungkan oleh penyeimbangan kelas. Sementara grafik perbandingan AUC-ROC (dengan SMOTE) memperlihatkan bahwa seluruh model tetap mempertahankan AUC di atas 0,90 sehingga kemampuan pemisahan kelas tetap sangat baik, dengan LightGBM dan Random Forest sebagai yang tertinggi (sekitar 0,934), XGBoost sedikit di bawah namun masih di kisaran 0,93, dan Logistic Regression sekitar 0,90 meskipun sedikit menurun dibandingkan skenario tanpa SMOTE. Jika grafik akurasi dan AUC-ROC ini dibandingkan dengan kondisi tanpa SMOTE, tampak bahwa SMOTE memberikan keuntungan paling besar pada model ensemble berbasis pohon (Random Forest dan LightGBM) karena keduanya tidak hanya meningkatkan akurasi dan F1-score, tetapi juga mampu mempertahankan bahkan sedikit menaikkan nilai AUC-ROC; oleh karena itu penelitian ini memilih Random Forest dengan SMOTE sebagai model terbaik karena memberikan kombinasi paling seimbang antara akurasi tinggi, AUC-ROC tinggi, dan penurunan signifikan pada jumlah *false negative*.

3.7 Pemilihan Model Terbaik

Hasil dua skenario eksperimen, yaitu klasifikasi penyakit jantung tanpa penanganan ketidakseimbangan kelas dan dengan penerapan SMOTE. Secara umum, keempat algoritma yang diuji Logistic Regression, Random Forest, LightGBM, dan XGBoost mampu memberikan kinerja yang baik dengan akurasi di atas 0,83 dan nilai AUC-ROC di atas 0,90 pada kedua skenario. Hal ini menunjukkan bahwa kombinasi fitur klinis yang digunakan sudah cukup informatif untuk membedakan pasien sehat dan pasien yang berisiko penyakit jantung.

Pada skenario tanpa SMOTE, Logistic Regression menghasilkan akurasi dan F1-score sedikit lebih tinggi dibandingkan LightGBM, XGBoost, dan Random Forest, sementara Random Forest memperoleh nilai AUC-ROC tertinggi. Pola ini memperlihatkan adanya trade-off: model linier seperti Logistic Regression cenderung lebih stabil dan seimbang antara kelas positif dan negatif, sedangkan Random Forest lebih unggul dalam memisahkan skor probabilitas antara dua kelas sehingga kurva ROC-nya lebih baik [32]. Confusion matrix juga menunjukkan bahwa Logistic Regression memiliki jumlah false negative yang relatif lebih rendah dibanding Random Forest, sehingga lebih sedikit pasien sakit yang luput terdeteksi, meskipun Random Forest memiliki true negative lebih banyak [9], [31].

Setelah diterapkan SMOTE pada data latih, terjadi perubahan yang cukup jelas. Secara umum, recall kelas positif meningkat pada semua model, yang berarti model menjadi lebih sensitif dalam mendeteksi pasien yang benar-benar sakit [10], [21]. Konsekuensinya, jumlah false positive sedikit bertambah, namun hal ini masih dapat diterima dalam konteks medis, karena salah mengklasifikasikan pasien sehat sebagai berisiko biasanya lebih dapat ditoleransi dibanding melewatkan pasien sakit. Nilai F1-score dan AUC-ROC cenderung naik atau setidaknya stabil, sehingga *oversampling* tidak hanya memperbaiki recall tetapi juga mempertahankan kemampuan diskriminatif model [9].

Pada skenario dengan SMOTE, Random Forest menjadi model dengan performa paling konsisten: akurasi tertinggi, F1-score tertinggi, dan AUC-ROC sangat kompetitif dengan LightGBM dan XGBoost [33]. Confusion matrix Random Forest dengan SMOTE memperlihatkan peningkatan jumlah true positive sekaligus menjaga true negative tetap tinggi, sehingga keseimbangan antara sensitivitas dan spesifisitas tercapai dengan baik. LightGBM dan XGBoost juga menunjukkan performa kuat, namun peningkatan akurasinya tidak sebesar Random Forest. Logistic Regression tetap kompetitif dan mudah diinterpretasikan, tetapi sedikit tertinggal dari model ensemble dalam hal F1-score pada data setelah diterapkan teknik SMOTE [33], [35].

Jika dibandingkan dengan penelitian sebelumnya yang umumnya hanya melatih model pada data imbang atau hanya mengejar akurasi, hasil ini memperlihatkan bahwa kombinasi dua langkah pra-pemrosesan data (imputasi, encoding, standardisasi dan penyeimbangan kelas dengan SMOTE diikuti tuning hyperparameter dan evaluasi menyeluruh (akurasi, precision, recall, F1, dan AUC-ROC) dapat menghasilkan model yang lebih andal untuk kebutuhan klinis. Secara khusus, penerapan SMOTE terbukti tidak sekadar menaikkan jumlah sampel minoritas secara mekanis, tetapi benar-benar membantu model mengenali pola pasien sakit dengan lebih baik tanpa mengorbankan kinerja global.

Berdasarkan keseluruhan analisis, Random Forest dengan SMOTE dapat direkomendasikan sebagai konfigurasi model terbaik dalam penelitian ini, karena memberikan kombinasi akurasi, precision, recall, F1-score, dan AUC-ROC yang tinggi serta distribusi prediksi yang paling seimbang antara kelas sehat dan sakit. Namun demikian, Logistic Regression tetap relevan ketika interpretabilitas menjadi prioritas, sedangkan LightGBM dan XGBoost cocok digunakan ketika sumber daya komputasi memadai dan diperlukan model dengan fleksibilitas tinggi.

4. KESIMPULAN

Berdasarkan hasil evaluasi kinerja model klasifikasi penyakit jantung, dapat disimpulkan bahwa penerapan teknik SMOTE berpengaruh signifikan terhadap peningkatan performa model, khususnya pada metrik recall, F1-score, dan AUC-ROC. Pada pengujian tanpa SMOTE, model Logistic Regression menunjukkan performa yang paling stabil dengan nilai *accuracy* 0,8478, *precision* 0,8775, *recall* 0,8431, *F1-score* 0,8600, dan AUC-ROC 0,9195. Sementara itu, Random Forest memiliki nilai *precision* (0,9080) dan AUC-ROC (0,9310) tertinggi, namun nilai recall yang lebih rendah (0,7745) mengindikasikan keterbatasan model dalam mendeteksi kelas minor. Model LightGBM dan XGBoost menunjukkan performa yang kompetitif, tetapi masih berada sedikit di bawah Logistic Regression dalam hal keseimbangan metrik

evaluasi. Setelah dilakukan penyeimbangan data menggunakan SMOTE, seluruh model menunjukkan peningkatan yang lebih konsisten, terutama pada nilai *recall*. Model Random Forest dengan SMOTE memberikan performa terbaik secara keseluruhan dengan *accuracy* 0,8695, *precision* 0,8900, *recall* 0,8725, *F1-score* 0,8811, dan AUC-ROC 0,9334, yang mencerminkan keseimbangan optimal antara ketepatan dan sensitivitas model. Penggunaan teknik SMOTE pada model Random Forest meningkatkan nilai recal secara signifikan. Hal ini menunjukan teknik SMOTE meningkatkan kemampuan model untuk mendeteksi anggota suatu kelas minor. Secara keseluruhan, hasil penelitian ini membuktikan bahwa SMOTE efektif dalam meningkatkan kinerja model klasifikasi penyakit jantung, dan Random Forest dengan SMOTE direkomendasikan sebagai model terbaik untuk mendukung deteksi dini penyakit jantung.

REFERENCES

[1] A. Sepharni, I. E. Hendrawan, C. Rozikin, and S. Karawang, "STRING (Satuan Tulisan Riset dan Inovasi Teknologi)." [Online]. Available: https://www.kaggle.com/fedesoriano/heart-

[2] A. Yogianto, A. Homaidi, and Z. Fatah, "Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung," *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 3, pp. 1720–1728, Jul. 2024, doi: 10.33379/gtech.v8i3.4495.

[3] A. Nurmasani and Y. Pristyanto, "ALGORITME STACKING UNTUK KLASIFIKASI PENYAKIT JANTUNG PADA DATASET IMBALANCED CLASS," 2021. [Online]. Available: www.ejournal.unib.ac.id/index.php/pseudocode

[4] I. M. Agus Oka Gunawan, I. D. A. Indah Saraswati, I. D. G. Riswana Agung, and I. P. Eka Putra, "Klasifikasi Penyakit Jantung Menggunakan Algoritma Decision Tree Series C4.5 Dengan Rapidminer," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 5, no. 2, pp. 73–83, Apr. 2023, doi: 10.47233/jteksis.v5i2.775.

[5] A. Sepharni, I. E. Hendrawan, C. Rozikin, and S. Karawang, "STRING (Satuan Tulisan Riset dan Inovasi Teknologi)." [Online]. Available: https://www.kaggle.com/fedesoriano/heart-

[6] F. Bukhari, S. Nurdiati, M. Khoirun Najib, and R. Nurul Amalia, "Deteksi Penyakit Jantung Menggunakan Metode Klasifikasi Decision Tree dan Regresi Logistik," 2023.

[7] D. Haganta Depari *et al.*, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *JURNAL INFORMATIK Edisi ke*, vol. 18, p. 2022.

[8] F. Handayani *et al.*, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network dalam Prediksi Penyakit Jantung".

[9] A. M. A. Rahim, I. Y. R. Pratiwi, and M. A. Fikri, "Klasifikasi Penyakit Jantung Menggunakan Metode Synthetic Minority Over-Sampling Technique dan Random Forest Classifier," *Indonesian Journal of Computer Science*, vol. 12, no. 5, 2023, doi: 10.33022/ijcs.v12i5.3413.

[10] A. F. N. Masruriyah, H. Y. Novita, C. E. Sukmawati, A. Ramadhan, S. Arif, and B. Dermawan, "Pengukuran Kinerja Model Klasifikasi dengan Data Oversampling pada Algoritma Supervised Learning untuk Penyakit Jantung," *Computer Science (CO-SCIENCE)*, vol. 4, no. 1, pp. 62–70, 2024, doi: 10.31294/coscience.v4i1.2389.

[11] D. S. Komputer and F. T. Informasi, "Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining," 2020. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[12] I. Tougui, A. Jilbab, and J. El Mhamdi, "Heart disease classification using data mining tools and machine learning techniques," *Health Technol (Berl)*, vol. 10, no. 5, pp. 1137–1144, Sep. 2020, doi: 10.1007/s12553-020-00438-1.

[13] Redwan Sony, "https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data."

[14] "https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data."

[15] Q. Al-Na'amneh *et al.*, "Autonomous Self-Adaptation in the Cloud: ML-Heal's Framework for Proactive Fault Detection and Recovery." [Online]. Available: www.ijacsa.thesai.org

[16] J. Yang and J. Guan, "A Heart Disease Prediction Model Based on Feature Optimization and Smote-Xgboost Algorithm," *Information*, vol. 13, no. 10, p. 475, 2022, doi: 10.3390/info13100475.

[17] H. A. Al-Alshaikh *et al.*, "Comprehensive evaluation and performance analysis of machine learning in heart disease prediction," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-58489-7.

[18] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," *Algorithms*, vol. 16, no. 2, p. 88, 2023, doi: 10.3390/a16020088.

[19] N. Alotaibi and M. Alzahrani, "Comparative Analysis of Machine Learning Algorithms and Data Mining Techniques for Predicting the Existence of Heart Disease." [Online]. Available: https://www.

[20] G. Husain *et al.*, "SMOTE vs. SMOTEENN: A Study on the Performance of Resampling Algorithms for Addressing Class Imbalance in Regression Models," *Algorithms*, vol. 18, no. 1, Jan. 2025, doi: 10.3390/a18010037.

[21] Y. A. Sir and A. H. H. Soepranoto, "Pendekatan Resampling Data Untuk Menangani Masalah Ketidakseimbangan Kelas," *Jurnal Komputer dan Informatika*, vol. 10, no. 1, pp. 31–38, Mar. 2022, doi: 10.35508/jicon.v10i1.6554.

[22] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 677–690, Jul. 2022, doi: 10.30812/matrik.v21i3.1726.

[23] Y. Han and I. Joe, "Enhancing Machine Learning Models Through PCA, SMOTE-ENN, and Stochastic Weighted Averaging," *Applied Sciences (Switzerland)*, vol. 14, no. 21, Nov. 2024, doi: 10.3390/app14219772.

[24] M. Ahmed, M. H. Sulaiman, M. M. Hassan, and T. Bhuiyan, "Predicting the Classification of Heart Failure Patients Using Optimized Machine Learning Algorithms," *IEEE Access*, vol. 13, pp. 30555–30569, 2025, doi: 10.1109/ACCESS.2025.3541069.

[25] R. Valarmathi and T. Sheela, "Heart Disease Prediction Using Hyper Parameter Optimization (HPO) Tuning," *Biomed Signal Process Control*, vol. 70, p. 103033, 2021, doi: 10.1016/j.bspc.2021.103033.

[26] E. R. Susanto and A. Eka Pranajaya, "Optimasi Random Forest untuk Prediksi Penyakit Jantung Menggunakan SMOTEENN dan Grid Search," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 7, pp. 1965–1979, Jul. 2025, doi: 10.52436/1.jpti.855.

[27] K. Sumwiza, C. Twizere, G. Rushingabigwi, P. Bakunzibake, and P. Bamurigire, "Enhanced Cardiovascular Disease Prediction Model Using Random Forest Algorithm," *Inform Med Unlocked*, vol. 41, p. 101316, 2023, doi: 10.1016/j.imu.2023.101316.

[28] M. Daviran, A. Maghsoudi, R. Ghezelbash, and B. Pradhan, "A new strategy for spatial predictive mapping of mineral prospectivity: Automated hyperparameter tuning of random forest 2 approach 3."
[29] M. G. El-Shafiey, A. Hagag, E. S. A. El-Dahshan, and M. A. Ismail, "A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest," *Multimed Tools Appl*, vol. 81, no. 13, pp. 18155–18179, May 2022, doi: 10.1007/s11042-022-12425-x.
[30] N. Chandrasekhar and S. Peddakrishna, "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization," *Processes*, vol. 11, no. 4, p. 1210, 2023, doi: 10.3390/pr11041210.
[31] T. Gori and A. Hestiningtyas, "Optimasi Pemilihan Fitur untuk Prediksi Penyakit Jantung Menggunakan Algoritma Genetika dan Random Forest," *Indonesian Journal of Computer Science*, vol. 13, no. 5, 2024, doi: 10.33022/ijcs.v13i5.4214.
[32] N. H. Alfajr and S. Defiyanti, "Prediksi Penyakit Jantung Menggunakan Metode Random Forest dan Penerapan Principal Component Analysis (PCA)," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, 2024, doi: 10.23960/jitet.v12i3S1.5055.
[33] V. dan T. S. dan V. S. P. Ramalingamsakthivelan N. M. K. dan Silambarasan, "Heart Disease Risk Assessment by Using LightGBM Technique," *International Journal for Multidisciplinary Research*, vol. 5, no. 2, 2023, [Online]. Available: https://www.ijfmr.com/papers/2023/2/2620.pdf
[34] K. V. Tompra, G. Papageorgiou, and C. Tjortjis, "Strategic Machine Learning Optimization for Cardiovascular Disease Prediction and High-Risk Patient Identification," *Algorithms*, vol. 17, no. 5, May 2024, doi: 10.3390/a17050178.
[35] J. Yang Jian dan Guan, "A Heart Disease Prediction Model Based on Feature Optimization and Smote-Xgboost Algorithm," *Information*, vol. 13, no. 10, p. 475, 2022, doi: 10.3390/info13100475.