Klasifikasi Keberadaan Penyakit Jantung Menggunakan XGBoost

Jati Tepatasa Bagastakwa
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
jtbagastakwa@student.telkomuniversity.ac.id

Jondri
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
jondri@telkomuniversity.ac.id

Indwiarti
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
indwiarti@telkomuniversity.ac.id

Abstrak — Penyakit kardiovaskular merupakan salah satu penyebab utama kematian di seluruh dunia dan biasanya memerlukan prosedur diagnosis medis yang kompleks. Penelitian ini bertujuan untuk membangun model klasifikasi machine learning menggunakan metode Extreme Gradient Boosting (XGBoost) untuk mengidentifikasi potensi penyakit kardiovaskular berdasarkan data pemeriksaan medis. Kumpulan data yang digunakan memiliki fitur-fitur seperti usia, tekanan darah, kadar kolesterol, status merokok, dan lain lain. Tahapan penelitian meliputi normalisasi dan pembersihan data, analisis data eksploratori (eksplorasi data), pelatihan model XGBoost, optimasi hyperparameter dengan library Optuna, dan evaluasi model berbasis metrik akurasi, presisi, recall, f1-score, dan AUC-ROC. Penelitian ini menghasillan model dengan nilai metrik akurasi sebesar 0.88, presisi sebesar 0.81, recall sebesar 0.96, f1-score sebesar 0.88, dan AUC-ROC sebesar 0.94. Nilai-nilai metrik dari model ini menunjukkan bahwa model bekerja dengan baik dalam mengklasifikasi keberadaan penyakit kardiovaskular. Penelitian ini juga memaparkan tingkat pengaruh dari faktor-faktor risiko terhadap prediksi model. Telah didapatkan bahwa secara berurutan faktor-faktor risiko dari yang paling berpengaruh di urutan tiga teratas, yaitu thal atau Thallium Stress Test (sebuah tes pencitraan untuk melihat aliran darah ke otot jantung di mana nilai dapat merepresentasikan kondisi: 3 = normal, 6 = cacat tetap, 7 = cacat reversibel), ca atau jumlah pembuluh darah utama (0-3) yang terwarnai (tersumbat) berdasarkan tes fluoroskopi, dan cp atau jenis nyeri dada yang dialami pasien (misalnya, angina tipikal, angina atipikal, dll).

Kata kunci— kesehatan, jantung, kardiovaskular, machine learning, xgboost.

I. PENDAHULUAN

A. Latar Belakang

Penyakit jantung adalah salah satu penyebab utama kematian di seluruh dunia. Penyakit kardiovaskular adalah penyakit yang disebabkan oleh gangguan fungsi jantung dan pembuluh darah [1]. Penyakit ini dapat memengaruhi satu atau banyak bagian jantung dan/atau pembuluh darah. Untuk mengetahui ada tidaknya penyakit ini, seorang individu perlu melakukan pemeriksaan medis. Setelah pemeriksaan, berdasarkan data yang tersedia, dokter akan memeriksa ada tidaknya potensi penyakit kardiovaskular pada pasien/individu tersebut. Berdasarkan tindakan tersebut, penelitian ini dilaksanakan untuk menunjang pekerjaan dokter dalam memeriksa ada tidaknya penyakit kardiovaskular dari hasil data pemeriksaan yang telah dilakukan.

Metode klasifikasi yang digunakan pada penelitian ini adalah dengan algoritma Extreme Gradient Boosting (XGBoost). Sebelumnya, penelitian klasifikasi penyakit kardiovaskular sudah pernah dilakukan dengan berbagai macam metode algoritma [2][3]. Beberapa metode yang umum digunakan dalam klasifikasi penyakit kardiovaskular adalah Support Vector Machine (SVM), Random Forest, dan Extreme Gradient Boosting (XGBoost) [4]. Penelitian oleh Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Nusa Mandiri, Jakarta, Indonesia menunjukkan bahwa metode SVM, XGBoost, dan Random Forest cukup efektif dalam mendeteksi penyakit jantung [3]. Istilah "efektif" dalam konteks ini merujuk pada nilai metrik evaluasi seperti akurasi, presisi, recall, dan f1-score yang tinggi pada penelitian sebelumnya. Pada penelitian kali ini, diterapkan optimasi hyperparameter secara otomatis menggunakan library Optuna untuk mendapatkan model yang optimal. Optuna menggunakan pendekatan optimasi Bayesian berbasis Tree-structured Parzen Estimator (TPE) untuk mengeksplorasi ruang parameter model secara efisien.

Meskipun algoritma machine learning seperti XGBoost memiliki performa tinggi, penerapannya pada data medis menghadapi tantangan kualitas data. Data rekam 1 medis seringkali tidak lengkap (missing values), yang jika tidak ditangani dengan metode imputasi yang tepat (seperti Modus atau KNN), dapat menurunkan akurasi diagnosis. Selain itu, data medis cenderung memiliki ketidakseimbangan kelas (imbalanced class), di mana jumlah pasien sehat lebih banyak daripada pasien sakit. Hal ini berisiko membuat model bias ke kelas mayoritas dan gagal mendeteksi pasien yang sebenarnya sakit (rendahnya nilai Recall).

Tantangan lain adalah sifat black-box dari algoritma ensemble yang sulit diinterpretasi. Dalam dunia medis, mengetahui "mengapa" model memprediksi seseorang sakit sangatlah krusial. Oleh karena itu, pendekatan Explainable AI (XAI) menggunakan SHAP (SHapley Additive exPlanations) diperlukan untuk memberikan transparansi terhadap keputusan model.

B. Rumusan Masalah

i. Bagaimana pengaruh perbandingan metode imputasi Simple Imputer (Modus) dan KNN Imputer terhadap kinerja model XGBoost?
ii. Bagaimana efektivitas penanganan ketidakseimbangan data menggunakan metode scale_pos_weight dan SMOTE dalam meningkatkan nilai Recall model?
iii. Bagaimana interpretabilitas model dalam menjelaskan faktor risiko penyakit jantung menggunakan analisis SHAP?

C. Tujuan dan Manfaat

i. Menemukan pengaruh perbandingan metode imputasi Simple Imputer (Modus) dan KNN Imputer terhadap kinerja model XGBoost.
ii. Menemukan efektivitas penanganan ketidakseimbangan data menggunakan metode scale_pos_weight dan SMOTE dalam meningkatkan nilai Recall model.
iii. Menginterpretabilitas model dalam menjelaskan faktor risiko penyakit jantung menggunakan analisis SHAP.

D. Batasan Masalah

i. Kumpulan data yang digunakan dalam penelitian ini adalah kumpulan data publik untuk penyakit kardiovaskular, yaitu dari repositori UCI Machine Learning. Dataset yang digunakan berasal dari tahun 1989 yang ukurannya relatif kecil, yaitu 303 baris, sehingga generalisasi model pada populasi modern yang lebih beragam mungkin memerlukan validasi lebih lanjut.
ii. Penelitian ini hanya menggunakan fitur-fitur yang tersedia dalam data tanpa menambahkan atau mencari data eksternal lainnya.
iii. Model klasifikasi yang digunakan dalam penelitian ini terbatas pada penggunaan algoritma Extreme Gradient Boosting (XGBoost), tanpa membandingkan performansi dengan algoritma lain.
iv. Proses pemodelan dan evaluasi dilakukan dengan pemisahan data uji-latih acak (train-tes split) dan validasi silang (cross-validation) serta tidak memakai data real-time atau data baru dari masyarakat.

E. Metode Penelitian

Penelitian ini dilakukan dengan pendekatan kuantitatif eksperimental melalui empat tahapan utama:

i. Studi Literatur:
Mengkaji teori terkait penyakit kardiovaskular, algoritma XGBoost, teknik penanganan data medis, dan interpretasi model.
ii. Pra-pemrosesan Data:
Melakukan pembersihan data, normalisasi, dan seleksi metode imputasi terbaik (Modus vs KNN Imputer) untuk menangani nilai yang hilang.
iii. Pembangunan Model:
Melatih model XGBoost menggunakan data hasil imputasi terbaik, menerapkan teknik penanganan ketidakseimbangan kelas (scale_pos_weight dan SMOTE), serta optimasi hyperparameter otomatis dengan Optuna.
iv. Evaluasi dan Interpretasi:
Mengukur performa model menggunakan metrik akurasi, presisi, recall, f1-score, dan AUC-ROC, serta menganalisis transparansi prediksi model menggunakan metode SHAP (SHapley Additive exPlanations).

II. KAJIAN TEORI

Penelitian ini menganalisis data faktor risiko penyakit kardiovaskular (berdasarkan data pemeriksaan medis) menggunakan metode machine learning untuk membantu dokter dalam mengidentifikasi potensi penyakit kardiovaskular. Penelitian ini melakukan eksplorasi data (EDA), menyiapkan dataset untuk model, serta mengidentifikasi faktor yang paling berpengaruh dalam prediksi penyakit. Model nantinya dievaluasi berdasarkan akurasi, presisi, recall, f1 score dan AUC-ROC untuk menilai efektivitasnya dalam mengklasifikasikan potensi penyakit kardiovaskular.

A. Literature Review

Berikut adalah tabel literature review dari penelitian ini.

TABEL 2

No.	*Literature Review*		
	Judul	**Tujuan atau Ide Pokok**	***Review***
1	Klasifikasi Penyakit Jantung Menggunakan *Extreme Gradient Boosting* (XGBoost)[1].	1. Mengeksplorasi pemanfaatan algoritma *machine learning*, khususnya XGBoost, dalam klasifikasi penyakit jantung. 2. Mengukur efektivitas algoritma XGBoost dalam meningkatkan akurasi dan efisiensi sistem deteksi penyakit jantung.	Penelitian ini melakukan klasifikasi penyakit jantung menggunakan algoritma berbasis *machine learning*, yaitu *Extreme Gradient Boosting* atau XGBoost. Hasil pengujian pada penelitian ini menunjukkan bahwa XGBoost mampu mencapai performa yang baik dalam seluruh metrik evaluasi[1].
2	Perbandingan Metode *Machine Learning* Untuk Mendeteksi Penyakit Jantung[2].	Membandingkan performa beberapa algoritma *machine learning*, seperti XGBoost, *Random Forest*, MLP Classifier, SVM, KNN, Linear Regression, dan Gradient Boosting dalam memprediksi penyakit jantung.	Penelitian ini bertujuan untuk memprediksi penyakit jantung dengan menggunakan perbandingan dari algoritma m*achine learning*. Salah satunya dengan algoritma *Extreme Gradient Boosting* atau XGBoost. Dari penelitian ini didapatkan akurasi terbaik menggunakan algoritma XGBoost dengan akurasi mencapai 95,08%[2].
3	Perbandingan Metode	1. Mengembangkan model prediksi	Penelitian ini bertujuan untuk

No.	*Literature Review*		
	Judul	**Tujuan atau Ide Pokok**	***Review***
	Decision Tree Classifier dan XGBoost *Classifier* Dalam Memprediksi Penyakit Jantung[3].	penyakit jantung menggunakan algoritma *Decision Tree* dan XGBoost. 2. Membandingkan akurasi dan efektivitas kedua algoritma dalam memprediksi penyakit jantung.	memprediksi penyakit jantung dengan menggunakan perbandingan dari algoritma *Machine Learning* yaitu *Decision Tree* dan XGBoost. Hasil penelitian mengindikasikan bahwa algoritma XGBoost mencapai akurasi 93%, lebih unggul dibandingkan *Decision Tree* yang memperoleh akurasi 90%[3].

B. Kardiovaskular

Menurut KBBI (Kamus Besar Bahasa Indonesia), kardiovaskular adalah istilah dalam kedokteran yang berhubungan dengan jantung dan pembuluh darah. Sementara penyakit kardiovaskular berarti sekelompok penyakit yang memengaruhi jantung dan pembuluh darah [4]. Penyakit ini dapat memengaruhi satu atau banyak bagian jantung dan/atau pembuluh darah. Dalam penelitian ini, dilatih model *machine learning* XGBoost untuk klasifikasi biner ada atau tidaknya penyakit jantung dari dataset faktor risiko penyakit jantung. Tabel 2.2. adalah atribut pada dataset faktor risiko penyakit jantung yang digunakan untuk melatih model.

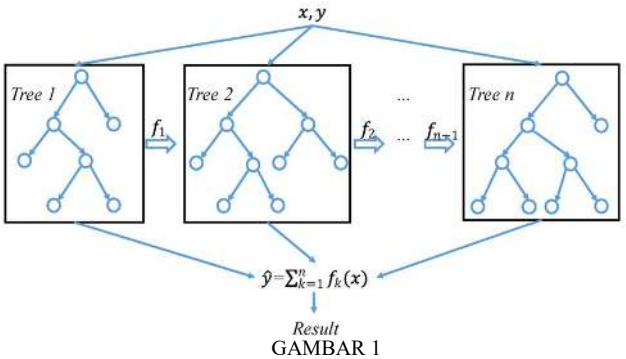
TABEL 3

Atribut	Deskripsi Dataset				
	Peran	**Tipe**	**Deskripsi**	**Unit**	**Missing Values**
age	Fitur	Integer	Usia pasien dalam tahun	years	Tidak ada
sex	Fitur	Categorical	Jenis kelamin pasien (1 = pria; 0 = wanita)		Tidak ada
cp	Fitur	Categorical	Jenis nyeri dada yang dialami pasien (misalnya, angina tipikal, angina atipikal, dll.)		Tidak ada
trestbps	Fitur	Integer	Tekanan darah saat istirahat (diukur saat pasien masuk rumah sakit)	mm Hg	Tidak ada
chol	Fitur	Integer	Kadar kolesterol serum (total kolesterol dalam darah)	mg/dl	Tidak ada
fbs	Fitur	Categorical	Gula darah puasa > 120 mg/dl. (1 = ya; 0 = tidak)		Tidak ada
restecg	Fitur	Categorical	Hasil elektrokardiogram (EKG) saat istirahat, yang menunjukkan kelainan irama jantung.		Tidak ada
thalach	Fitur	Integer	Detak jantung maksimum yang tercapai saat uji latih jantung (stress test)		Tidak ada
exang	Fitur	Categorical	Angina (nyeri dada) yang dipicu oleh olahraga. (1 = ya; 0 = tidak)		Tidak ada
oldpeak	Fitur	Integer	Depresi segmen ST pada EKG yang disebabkan oleh olahraga relatif terhadap kondisi istirahat. Mengindikasikan iskemia (kurangnya aliran darah).		Tidak ada
slope	Fitur	Categorical	Kemiringan (slope) segmen ST pada puncak latihan, memberikan informasi tentang respons jantung.		Tidak ada
ca	Fitur	Integer	Jumlah pembuluh darah utama (0-3) yang terwarnai (tersumbat) berdasarkan tes fluoroskopi.		Ada
thal	Fitur	Categorical	Hasil tes stres talium (Thallium Stress Test), sebuah tes pencitraan untuk melihat aliran darah ke otot jantung (di mana nilai dapat merepresentasikan kondisi: 3 = normal, 6 = cacat tetap, 7 = cacat reversibel).		Ada

C. *Extreme Gradient Boosting* (XGBoost)

Extreme Gradient Boosting (XGBoost) adalah sebuah algoritma *machine learning* yang sangat populer dan efektif dalam analisis data, khususnya dalam tugas-tugas seperti klasifikasi, regresi, dan peringkatan. Algoritma ini dikembangkan oleh Tianqi Chen pada tahun 2014 dan menjadi sangat populer karena kecepatan, efisiensi, dan kemampuannya menghasilkan prediksi yang akurat [5]. XGBoost adalah metode *ensemble* yang menggabungkan banyak pohon keputusan (*decision tree*) dengan menggunakan pendekatan boosting. Performa model ini

sangat bergantung pada penyetelan *hyperparameter* saat menggunakannya. Beberapa parameter yang biasanya dioptimalkan meliputi *max_depth*, yang mengontrol kompleksitas pohon, *min_child_weight*, dalam upaya mencegah *overfitting*, *subsample* dan *colsample_bytree*, yang menyediakan regularisasi stokastik, dan *learning_rate*, yang mengontrol laju pembelajaran. Selain itu, parameter regularisasi seperti *gamma*, *lambda* (regularisasi L2), dan *alpha* (regularisasi L1) digunakan untuk mengurangi kompleksitas model. Dalam penelitian ini, optimasi *hyperparameter* dilakukan menggunakan Optuna, pustaka optimasi cepat dan otomatis berdasarkan optimasi Bayesian.



GAMBAR 1

D. Optuna

Optuna adalah suatu pustaka (*library*) *open-source* untuk melakukan optimasi *hyperparameter* secara efisien dan otomatis [6]. Optuna sangat sering dipakai untuk melakukan *tuning* model-model *machine learning* (ML) seperti XGBoost [7][8], LightGBM, *Random Forest*, *Neural Network*, dan sebagainya. Optuna menggunakan pendekatan optimasi Bayesian berbasis *Tree-structured Parzen Estimator* (TPE) untuk mengeksplorasi ruang parameter model secara efisien.

E. SHAP *(SHapley Additive exPlanations)*

SHAP (SHapley Additive exPlanations) adalah sebuah pendekatan untuk menginterpretasikan output dari model machine learning. Metode ini diperkenalkan oleh Lundberg dan Lee pada tahun 2017 dengan tujuan untuk menjelaskan prediksi dari model yang kompleks (sering disebut sebagai black-box), seperti ensemble tree (XGBoost, Random Forest) atau neural networks, agar lebih transparan dan dapat dipahami manusia. SHAP menghubungkan alokasi kredit optimal dengan penjelasan lokal menggunakan Shapley values yang berasal dari teori permainan kooperatif (cooperative game theory) [10].

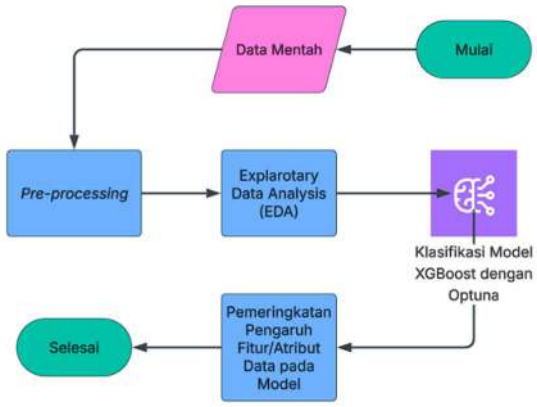
Konsep dasar SHAP mengibaratkan sebuah prediksi model sebagai hasil dari sebuah permainan, di mana setiap fitur atau atribut data dianggap sebagai "pemain" yang berkontribusi terhadap hasil tersebut. SHAP menghitung kontribusi marginal (marginal contribution) dari setiap fitur terhadap hasil prediksi akhir. Kontribusi marginal ini dihitung dengan melihat selisih prediksi ketika sebuah fitur diikutsertakan dalam model dibandingkan ketika fitur tersebut tidak diikutsertakan, yang dirata-ratakan atas semua kemungkinan kombinasi fitur lainnya [10].

Secara matematis, nilai SHAP untuk fitur tertentu merepresentasikan seberapa jauh fitur tersebut menggeser prediksi model dari nilai rata-rata dasar (baseline). Nilai SHAP yang positif mengindikasikan bahwa fitur tersebut mendorong prediksi ke arah yang lebih tinggi (misalnya, meningkatkan probabilitas terdeteksi penyakit), sedangkan nilai SHAP negatif mengindikasikan fitur tersebut mendorong prediksi ke arah yang lebih rendah. Keunggulan utama SHAP dibandingkan metode interpretasi lainnya adalah sifat aditif dan konsistensinya, yang menjamin bahwa jika suatu model mengubah ketergantungannya pada suatu fitur, nilai atribusi fitur tersebut tidak akan berkurang [10].

III. METODE

Penelitian ini bertujuan untuk membantu dokter mendeteksi kemungkinan penyakit kardiovaskular pada seseorang menggunakan data pemeriksaan medis melalui penerapan model klasifikasi XGBoost. Penelitian terdiri dari tiga fase utama, yaitu pra-pemrosesan data, identifikasi faktor yang paling memengaruhi penyakit, dan evaluasi kinerja model berkenaan dengan metrik akurasi, presisi, recall, f1-score dan AUC-ROC. XGBoost digunakan karena kemahirannya, terbukti dalam penelitian sebelumnya dibandingkan dengan kinerja algoritma lain seperti SVM dan Random Forest.

A. Diagram Pemodelan



GAMBAR 2

Berikut adalah penjelasan diagram pemodelan:

i. Data Mentah

Data mentah berbentuk tabular yang berisikan data faktor risiko penyakit kardiovaskular. Data memiliki 13 kolom/atribut dan terdiri dari 303 baris berisikan data pasien. Data di ambil dari https://archive.ics.uci.edu/dataset/45/heart+disease Cleveland subset [9].

ii. Pre-processing

Sebelum dipakai data dibersihkan/dirapikan. Setelah data bersih, data dinormalisasikan. Normalisasi data dilakukan dengan MinMaxScaler() dari library sklearn.preprocessing. Penanganan missing values dilakukan dengan membandingkan dua metode imputasi, yaitu Simple Imputer (modus) dan KNN Imputer. Metode imputasi yang terbukti memberikan performa terbaik akan dipilih untuk

memproses data yang akan digunakan pada tahap klasifikasi selanjutnya.

iii. Exploratory Data Analysis (EDA)

Pada tahap ini, EDA yang dilakukan berupa pengecekkan statistika deskriptif dari data, distribusi nilai-nilai pada data, dan nilai korelasi antar atribut/kolom pada data.

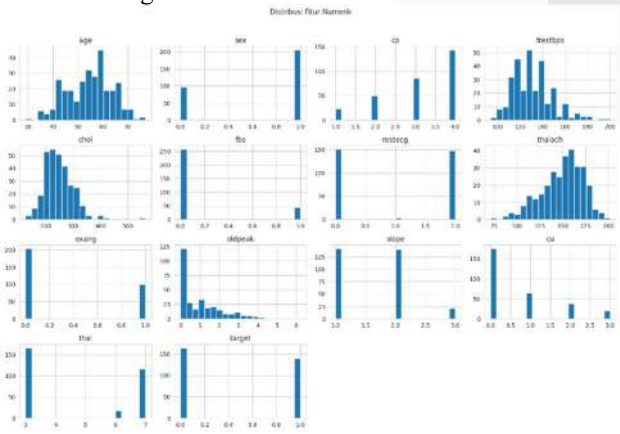
a. Statistika Deskriptif Data

TABEL 4

Fitur (Atribut)	count	mean	std	min	25%	50%	75%	max
Age	303	54.43894	9.038662	29	48	56	61	77
Sex	303	0.679868	0.467299	0	0	1	1	1
Cp	303	3.158416	0.960126	1	3	3	4	4
Trestbps	303	131.6898	17.59975	94	120	130	140	200
Chol	303	246.6931	51.77692	126	211	241	275	564
Fbs	303	0.148515	0.356198	0	0	0	0	1
Restecg	303	0.990099	0.994971	0	0	1	2	2
Thalach	303	149.6073	22.875	71	133.5	153	166	202
Exang	303	0.326733	0.469794	0	0	0	1	1
Oldpeak	303	1.039604	1.161075	0	0	0.8	1.6	6.2
Slope	303	1.60066	0.616226	1	1	2	2	3
Ca	299	0.672241	0.937438	0	0	0	0	3
Thal	301	4.734219	1.939706	3	3	3	7	7
Target	303	0.458746	0.49912	0	0	0	1	1

b. Distribusi Data

Gambar 3.2. merupakan distribusi data dari fitur kontinu dan kategoris. Variabel kontinu seperti usia dan trestbps digambarkan dengan distribusi yang relatif normal, sementara oldpeak sangat miring. Salah satu poin utama adalah ketidakseimbangan yang cukup besar pada beberapa fitur kategoris (misalnya sex dan fbs) yang dapat menyebabkan model pembelajaran mesin yang bias. Namun demikian, variabel target bersifat biner dan tampaknya cukup seimbang, yang secara positif menunjukkan bahwa variabel tersebut dapat digunakan untuk tugas klasifikasi.



GAMBAR 3

c. Distribusi Data

Pada peta korelasi Spearman ini, dapat dilihat bahwa telah berhasil ditemukan prediktor utama dari variabel target. Atribut yang memiliki korelasi positif tertinggi dengan variabel target adalah thal (0.52), ca (0.49), cp (0.46), exang (0.42), dan oldpeak (0.41). Di sisi lain, thalach menonjol dengan korelasi negatif tertinggi (-0.43). Fitur seperti fbs tidak memiliki korelasi yang signifikan (0.00), yang menunjukkan bahwa mereka tidak dapat membuat prediksi apa pun tentang target dalam kumpulan data ini tanpa bantuan variabel lain.

GAMBAR 4

iv. Klasifikasi dengan Model XGBoost

Pada tahap ini, dilakukan pembangunan model klasifikasi menggunakan algoritma Extreme Gradient Boosting (XGBoost). Proses pelatihan model ini menggunakan dataset bersih yang telah diproses menggunakan metode imputasi terbaik dari tahap pre-processing. Selanjutnya, pada tahap ini juga diterapkan teknik penanganan ketidakseimbangan kelas (class imbalance handling) untuk memastikan model tidak bias. Model kemudian dioptimasi secara otomatis menggunakan Optuna untuk mendapatkan hyperparameter terbaik. Terakhir, kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, f1-score, dan AUC-ROC.

v. Pemeringkatan Pengaruh Fitur/Atribut Data pada Model

Pada tahap ini, dilakukan pemeringkatan pengaruh fitur/atribut data pada model.

IV. HASIL DAN PEMBAHASAN

I.Deskripsi Percobaan

i. Informasi Data Percobaan

a. Jumlah baris awal adalah 303 baris.

b. Jumlah nilai yang hilang (*missing values*) per kolom:

1. Age : 0
2. Sex : 0
3. cp : 0
4. trestbps : 0
5. chol : 0

6. fbs : 0
7. restecg : 0
8. thalach : 0
9. exang : 0
10. oldpeak : 0
11. slope : 0
12. ca : 4
13. thal : 2
14. target : 0

c. Hasil Transformasi Target:
1. 0 : 164
2. 1 (wakili 1-4) : 139

d. Pembagian Pelatihan Data:
1. 242 baris untuk data latih.
2. 61 baris untuk data uji.

ii. Alur Eksperimen Percobaan

Alur eksperimen percobaan dalam penelitian ini dilakukan dengan skenario bertingkat (sequential), di mana hasil optimal dari satu skenario menjadi dasar bagi skenario berikutnya.

a. Skenario 1 (Imputasi):
Bertujuan mencari metode imputasi terbaik. Metode terpilih akan diterapkan untuk menghasilkan dataset final.

b. Skenario 2 (Penanganan Ketidakseimbangan):
Menggunakan dataset hasil Skenario 1, penelitian dilanjutkan dengan menguji teknik penanganan ketidakseimbangan kelas (imbalance handling).

c. Skenario 3 (Analisis Fitur):
Model terbaik dari Skenario 2 kemudian dianalisis transparansinya menggunakan metode Gain, Permutation Importance, dan SHAP.

iii. Skenario 1: Perbandingan Metode Imputasi

Pada skenario ini akan dilakukan analisis perbandingan skenario imputasi missing values dengan nilai modus dan nilai dari KNN Imputer.

iv. Skenario 2: Perbandingan Penanganan Ketidakseimbangan Kelas

Pada skenario ini akan dilakukan analisis perbandingan penanganan ketidakseimbangan kelas dengan:

a. Kondisi A: Baseline (Tanpa Penanganan Khusus)
b. Kondisi B: Menggunakan 'scale_pos_weight'
c. Kondisi C: Menggunakan SMOTE (Synthetic Minority Over-sampling Technique)

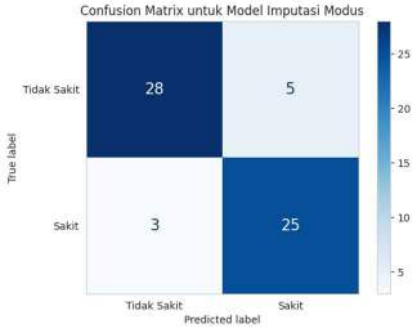
v. Skenario 3: Perbandingan Analisis Faktor Paling Berpengaruh (Feature Importance)

Pada skenario ini akan dilakukan analisis perbandingan faktor paling berpengaruh dengan metode gain (kepentingan fitur bawaan XGBoost), permutation importance, dan SHAP.

II.Hasil dan Analisis Percobaan

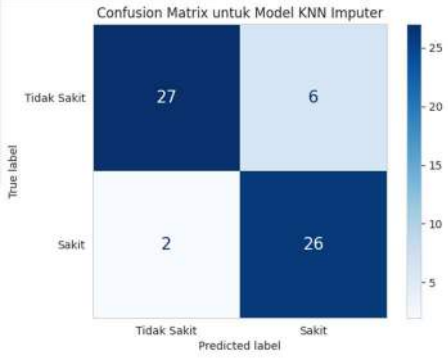
i. Skenario 1

a. Confusion Matrix untuk Imputasi Modus



GAMBAR 5

b. Confusion Matrix untuk KNN Imputer



GAMBAR 6

c. Analisis Perbandingan Skenario Imputasi

TABEL 5

Metrik	Imputasi Modus	KNN *Imputer*
n_estimators	139.000000	414.000000
learning_rate	0.019708	0.017123
max_depth	8.000000	7.000000
subsample	0.781705	0.800657
colsample_bytree	0.915576	0.971363
gamma	0.001241	0.001592

TABEL 6

Metrik	Imputasi Modus	KNN *Imputer*
Akurasi	0.868852	0.868852
Presisi	0.833333	0.812500
Recall	0.892857	0.928571
F1-Score	0.862069	0.866667
AUC-ROC	0.950216	0.936147

Keputusan: Model dengan KNN Imputer terpilih sebagai model terbaik. Meskipun akurasi kedua metode sama, model dengan KNN Imputer dipilih karena nilai Recall (0.928) dan F1-Score (0.866) yang sedikit lebih tinggi, yang lebih krusial dalam konteks medis (untuk meminimalkan risiko false negative).

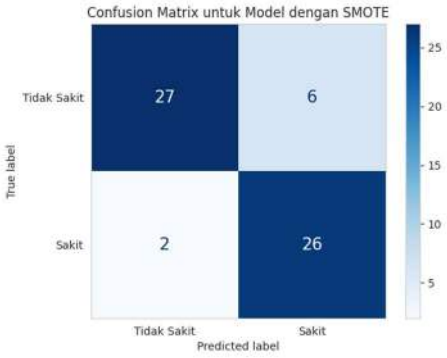
ii. Skenario 2

a. Kondisi A: Baseline (Tanpa Penanganan Khusus)
Model Baseline (dari data seimbang) sudah dievaluasi. Hasilnya akan digunakan sebagai pembanding.

b. Kondisi B: Menggunakan '*scale_pos_weight*'
   1. 'scale_pos_weight' adalah parameter XGBoost yang memberikan bobot lebih pada kelas minoritas. Ini "memberitahu" model untuk lebih memperhatikan kesalahan pada kelas positif. Nilainya dihitung sebagai: (jumlah sampel negatif / jumlah sampel positif).
   2. Nilai scale_pos_weight yang dihitung: 1.18
   3. Akurasi: 0.8852
   4. Presisi: 0.8182
   5. Recall (Sensitivitas): 0.9643
   6. F1-Score: 0.8852
   7. AUC-ROC: 0.9405
   8. Confusion Matrix untuk Model dengan scale_pos_weight

GAMBAR 7

c. Kondisi C: Menggunakan SMOTE (Synthetic Minority Over-sampling Technique)
   1. SMOTE bekerja dengan membuat sampel 'sintetis' baru dari kelas minoritas. Ini dilakukan dengan mengambil sampel dari kelas minoritas dan membuat data baru di antara sampel tersebut. SMOTE hanya diterapkan pada data latih untuk mencegah kebocoran data. Di sini menggunakan ImbPipeline untuk memastikan SMOTE hanya berjalan pada data latih di dalam validasi silang.
   2. Akurasi: 0.8689
   3. Presisi: 0.8125
   4. Recall (Sensitivitas): 0.9286
   5. F1-Score: 0.8667
   6. AUC-ROC: 0.9426
   7. Confusion Matrix untuk Model dengan SMOTE



GAMBAR 8

d. Analisis Perbandingan Skenario Ketidakseimbangan Kelas

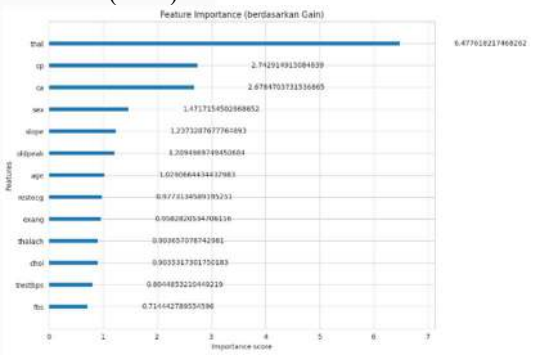
TABEL 7

Metrik	*Baseline* (Data Seimbang)	Dengan *scale_pos_weight*	Kenaikan Dengan *scale_pos_weight* (%)	Dengan SMOTE	Kenaikan Dengan SMOTE (%)
Akurasi	0.868852	0.885246	**+1.89%**	0.868852	**0.00 %**
Presisi	0.812500	0.818182	**+0.70%**	0.812500	**0.00 %**
Recall	0.928571	0.964286	**+3.85%**	0.928571	**0.00 %**
F1-Score	0.866667	0.885246	**+2.14%**	0.866667	**0.00 %**
AUC-ROC	0.936147	0.940476	**+0.46%**	0.942641	**+0.69 %**

Keputusan: Tabel 4.3 menunjukkan bahwa metode scale_pos_weight berhasil meningkatkan Recall sebesar 3.85% dibandingkan baseline. Peningkatan ini sangat signifikan dalam konteks medis untuk mengurangi risiko pasien sakit yang tidak terdeteksi

iii. Skenario 3

a. Metode A: Analisis Kepentingan Fitur Bawaan XGBoost (Gain)



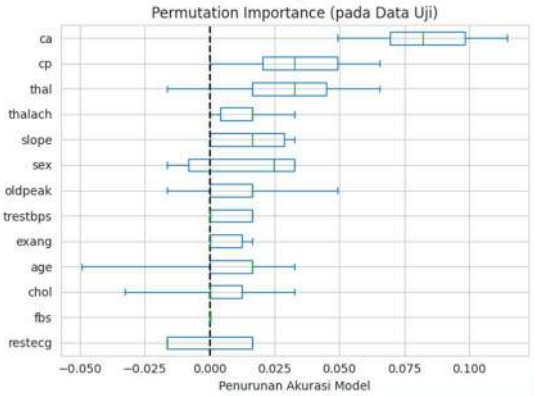
GAMBAR 9

Analisis:

1. Fitur thal sejauh ini adalah yang paling krusial: dengan skor kepentingan sekitar 6,48, lebih dari dua kali lipat skor fitur terpenting berikutnya.
2. cp (jenis nyeri dada) dan ca (jumlah pembuluh darah utama yang diwarnai dengan fluoroskopi) dianggap sebagai fitur terpenting kedua dan ketiga, dengan skor masing-masing 2,74 dan 2,68.
3. Terdapat penurunan tajam skor kepentingan setelah tiga fitur teratas, yang menunjukkan bahwa model sangat bergantung pada fitur-fitur ini untuk mencapai keputusannya.

4. fbs dan trestbps (tekanan darah istirahat) berada di posisi terbawah peringkat sebagai fitur yang paling tidak signifikan

b. Metode B: Permutation Importance

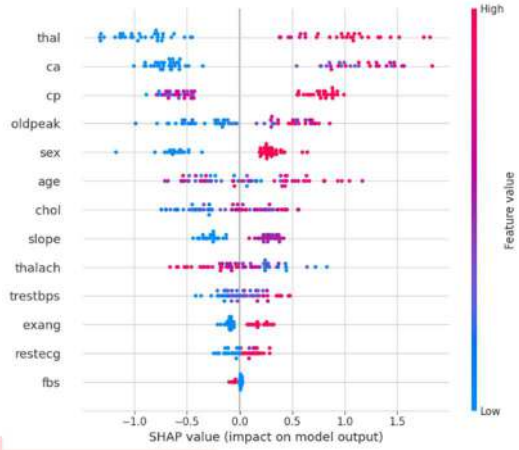


GAMBAR 10

Analisis:

1. Fitur Utama: Metode ini menyoroti ca, cp, dan thal sebagai tiga fitur teratas dengan efek terbesar. Plotnya juga sesuai dengan kesimpulan ini, tetapi urutannya telah berubah; ca sekarang berada di peringkat pertama di antara ketiganya, yang dapat menjadi alasan peringkat metode ini berbeda. Hal ini pada dasarnya mencerminkan skenario di mana kepentingan permutasi dilakukan pada set pengujian dan kontribusi fitur terhadap kemampuan model untuk menggeneralisasi ke data baru yang belum pernah dilihat sebelumnya sedang diukur.
2. Fitur dengan Kepentingan Rendah: Variabel chol, fbs, dan restecg terletak di dekat garis horizontal yang mewakili nol. Nilai-nilai mereka hampir tidak berdampak pada akurasi; pengacakan (permutation) terhadap fitur-fitur tersebut hampir tidak menurunkan kinerja model.
3. Kepentingan Negatif: Skor signifikansi negatif untuk `restecg` menunjukkan bahwa karakteristik ini cenderung hanya menawarkan tebak-tebakan dalam prediksi selama pengujian atau bahkan noise. Penghapusan informasi fitur ini (melalui pengaturan nilai yang acak) secara tidak sengaja dapat menyebabkan peningkatan kinerja model.

c. Metode C: Analisis SHAP



GAMBAR 11

Plot SHAP (SHapley Additive exPlanations) ini menggambarkan signifikansi dan pengaruh setiap fitur terhadap prediksi model. Fitur-fitur tersebut diurutkan berdasarkan tingkat kepentingannya, dengan thal, ca, dan cp menjadi tiga fitur paling berpengaruh.

1. Cara Membaca Plot
   a. Sumbu Y (Atribut): Menampilkan daftar atribut berurutan dari yang paling penting hingga yang paling tidak penting. "Penting" di sini berarti memiliki dampak nilai rata-rata absolut terbesar pada prediksi.
   b. Sumbu X (SHAP Value): Ini adalah dampak pada output model. Jika nilai positif (> 0), maka mendorong prediksi model ke arah yang lebih tinggi (probabilitas lebih tinggi untuk "terkena penyakit jantung"). Jika nilai negatif (< 0), maka mendorong prediksi model ke arah yang lebih rendah (probabilitas lebih rendah, alias "sehat").
   c. Warna (Nilai Fitur): Warna suatu titik menunjukkan nilai asli dari fitur tersebut untuk satu data (pasien). Tingkat skala warna biru (Low) sampai merah (High) menunjukkan apakah nilai fitur tersebut rendah atau tinggi. Contoh, pada fitur 'age', titik biru menggambarkan pasien dengan usia muda, dan titik merah menggambarkan pasien dengan usia tua.
2. Analisis Setiap Atribut
   a. thal (Tes Stres Talium):
      i. Tingkat Kepentingan: Atribut paling penting
      ii. Penjelasan: Atribut ini memiliki kontribusi positif yang sangat kuat terhadap prediksi penyakit jantung. Titik-titik berwarna merah

terpisah jelas dengan titik-titik berwarna biru. Nilai thal tinggi (merah) memiliki nilai SHAP positif kuat yang mendorong prediksi ke arah “terkena penyakit”. Sebaliknya, nilai thal rendah (biru) memiliki nilai SHAP negatif yang mendorong prediksi ke arah “sehat”.

b. ca (Jumlah Pembuluh Darah Utama yang Terdeteksi):
   i. Tingkat Kepentingan: Atribut terpenting kedua.
   ii. Penjelasan: Jumlah pembuluh darah utama yang terdeteksi bermasalah oleh fluoroskopi adalah indikator risiko yang sangat kuat. Semakin banyak pembuluh darah yang bermasalah (nilai ca tinggi), semakin tinggi model memprediksi risiko penyakit jantung. Jumlah pembuluh darah utama yang terdeteksi memiliki masalah (nilai ca tinggi, ditandai dengan titik merah) secara konsisten memberikan nilai SHAP positif, yang kuat sekali mendukung prediksi risiko. Nilai ca rendah (biru) menstimulasi prediksi ke arah 'sehat'.
c. cp (Tipe Nyeri Dada):
   i. Tingkat Kepentingan: Atribut terpenting ketiga.
   ii. Penjelasan: Model mempelajari bahwa tipe nyeri dada tertentu merupakan prediktor kuat untuk risiko keberadaan penyakit. Ini mengindikasikan bahwa tipe nyeri dada tertentu (dengan nilai kategorikal tinggi) adalah prediktor kuat penyakit.
d. oldpeak (Depresi ST Akibat Latihan):
   i. Tingkat Kepentingan: Atribut penting keempat.
   ii. Penjelasan: Menunjukkan hubungan yang jelas, di mana nilai oldpeak yang tinggi (merah) memiliki nilai SHAP positif yang tinggi. Nilai oldpeak yang rendah (biru) memiliki nilai SHAP negatif. Hal ini sejalan dengan pengetahuan medis.
e. sex (Jenis Kelamin):
   i. Tingkat Kepentingan: Atribut cukup penting.
   ii. Penjelasan: Nilai fitur 'sex' yang tinggi (merah, menyimbolkan 'pria') memiliki kemungkinan nilai SHAP positif, yang meningkatkan probabilitas prediksi penyakit. Nilai yang rendah (biru, menyimbolkan 'wanita') cenderung memiliki nilai SHAP negatif, yang menurunkan prediksi risiko.
f. age (Usia):
   i. Tingkat Kepentingan: Atribut cukup penting.
   ii. Penjelasan: Usia berpengaruh beragam. Akan tetapi, terdapat tendensi di mana usia lebih tua (merah) lebih banyak membentuk nilai SHAP positif (memperbesar risiko), dan usia lebih muda (biru) memiliki nilai SHAP negatif (meringkan risiko).
g. chol (Kolesterol):
   i. Tingkat Kepentingan: Atribut cukup penting, namun interpretasinya rumit.
   ii. Penjelasan: Plot menunjukkan tumpang tindihnya titik biru (rendah kolesterol) dan merah (tinggi kolesterol) di kedua ujungnya dari garis nol. Ini berarti tidak ada tren yang jelas; kedua kolesterol rendah dan

tinggi dapat mempengaruhi prediksi negatif atau positif, tergantung pada faktor lain.

h. slope (Kemiringan Segmen ST Puncak Latihan):
   i. Tingkat Kepentingan: Atribut cukup penting.
   ii. Penjelasan: Nilai slope yang rendah (biru) berkontribusi pada nilai SHAP positif (risiko lebih tinggi). Nilai slope yang tinggi (merah) berkontribusi pada nilai SHAP negatif (risiko lebih rendah).

i. thalach (Detak Jantung Maksimum yang Tercapai):
   i. Tingkat Kepentingan: Atribut cukup penting, dengan hubungan yang menarik.
   ii. Penjelasan: Ini berhubungan terbalik. Titik-titik merah (detak jantung maks tinggi) berada di sisi kiri (nilai SHAP negatif) dan titik-titik biru (detak jantung maks rendah) berada di sisi kanan (nilai SHAP positif). Detak jantung maksimum yang rendah saat (biru) adalah faktor risiko penyakit jantung. Sementara kemampuan mencapai detak jantung maksimum yang tinggi (merah) adalah tanda kebugaran yang menurunkan prediksi risiko penyakit jantung.

j. trestbps (Tekanan Darah Istirahat):
   i. Tingkat Kepentingan: Atribut kurang penting.
   ii. Penjelasan: Titik-titik (merah dan biru) hampir seluruhnya terkumpul di sekitar nilai SHAP 0, menunjukkan konsribusi yang kecil dan tidak konsisten terhadap prediksi.

k. exang (Angina Akibat Latihan):
   i. Tingkat Kepentingan: Atribut kurang penting.
   ii. Penjelasan: Masing-masing titik merah dan biru saling menumpuk secara sejenis, namun letaknya terpisah oleh garis nilai SHAP 0. Kedua jenis titik ini cukup dekat dengan SHAP 0, yang berarti memiliki dampak kecil pada model prediksi penyakit jantung.

l. restecg (Hasil EKG Istirahat):
   i. Tingkat Kepentingan: Atribut hampir tidak penting.
   ii. Penjelasan: Semua titik merah dan biru hampir menumpuk di tengah, dekat dengan nilai SHAP 0, yang berarti memiliki dampak kecil pada model prediksi penyakit jantung.

m. fbs (Gula Darah Puasa > 120 mg/dl)
   i. Tingkat Kepentingan: Atribut paling tidak penting.

a. Penjelasan: Semua titik merah dan biru menumpuk di tengah, sangat dekat dengan nilai SHAP 0, yang berarti memiliki dampak sangat kecil pada model prediksi penyakit jantung.

V. KESIMPULAN

Setelah serangkaian percobaan dan berbagai skenario dengan algoritma pembelajaran mesin (machine learning) XGBoost, diperoleh kesimpulan sebagai berikut:

1. Hasil perbandingan metode imputasi menunjukkan bahwa KNN Imputer lebih unggul dibandingkan Simple Imputer (Modus) terhadap kinerja model XGBoost. KNN Imputer terbukti memberikan kontribusi lebih baik dalam menjaga kualitas data, ditunjukkan dengan capaian nilai f1-score dan Recall yang lebih tinggi pada tahap pra-pemrosesan, sehingga metode ini dipilih sebagai metode imputasi terbaik.

2. Dalam hal efektivitas penanganan ketidakseimbangan data, metode scale_pos_weight terbukti lebih efektif dibandingkan SMOTE dalam meningkatkan nilai Recall model. Penerapan parameter scale_pos_weight berhasil mengoptimalkan sensitivitas model hingga mencapai skor Recall sebesar 0.9643. Angka ini sangat krusial dalam bidang medis untuk meminimalkan false negative (pasien sakit yang terprediksi sehat).

3. Analisis SHAP berhasil memberikan interpretabilitas yang transparan mengenai faktor risiko penyakit jantung. Berdasarkan analisis ini, fitur paling vital yang memengaruhi prediksi model secara berurutan adalah thal (tes stres talium), ca (jumlah pembuluh darah utama), dan cp (jenis nyeri dada). Secara spesifik, analisis SHAP menunjukkan bahwa nilai thal dan ca yang tinggi memiliki kontribusi positif yang kuat, yang mendorong prediksi model ke arah terdiagnosis penyakit.

Secara keseluruhan, model klasifikasi XGBoost yang telah difinalisasi dan dioptimasi melalui Optuna menghasilkan kinerja yang sangat baik dengan Akurasi sebesar 0.8852, Presisi sebesar 0.8182, Recall sebesar 0.9643, f1-score sebesar 0.8852, dan AUC-ROC sebesar 0.9405 pada data uji. Penelitian ini menyimpulkan bahwa XGBoost adalah metode yang andal untuk klasifikasi penyakit jantung, serta mampu memberikan pemahaman mendalam mengenai faktor risiko kardiovaskular melalui pendekatan Explainable AI.

REFERENSI

[1] M. Martiningsih and A. Haris, "RISIKO PENYAKIT KARDIOVASKULER PADA PESERTA PROGRAM PENGELOLAAN PENYAKIT KRONIS (PROLANIS) DI PUSKESMAS KOTA BIMA: KORELASINYA DENGAN ANKLE BRACHIAL INDEX DAN OBESITAS," *Jurnal Keperawatan Indonesia*, vol. 22, no. 3, pp. 200–208, Nov. 2019, doi: 10.7454/jki.v22i3.880.

[2] D. Bagus Darmawan, I. Rafli Azahwa, R. Wijaya Saputra, R. Septiadi, and P. Rosyani, "Klasifikasi Penyakit Jantung Menggunakan Extreme Gradient Boosting (XGBoost)," 2025. [Online]. Available: https://jurnalmahasiswa.com/index.php/jriin

[3] Y. Amelia, "PERBANDINGAN METODE MACHINE LEARNING UNTUK MENDETEKSI PENYAKIT JANTUNG," *IDEALIS : InDonEsiA journaL Information System*, vol. 6, no. 2, pp. 220–225, Jul. 2023, doi: 10.36080/idealis.v6i2.3043.

[4] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J. Schmid, S. Sandhu, K. Guppy, S. Lee, V. Froelicher, "International application of a new probability algorithm for the diagnosis of coronary artery disease," *Am J Cardiol*, vol. 64, no. 5, pp. 304–310, Aug. 1989, doi: 10.1016/0002-9149(89)90524-9.

[5] D. Mayang Pratiwi and L. Mufidah, *Perbandingan Metode Decision Tree Classifier dan XGBoost Classifier Dalam Memprediksi Penyakit Jantung*, vol. 4, no. 1. 2024.

[6] T. Chen and C. Guestrin, "XGBoost," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[7] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, New York, NY, USA: ACM, Jul. 2019, pp. 2623–2631. doi: 10.1145/3292500.3330701.

[8] A. Jain, A. Singh, and A. Doherey, "Prediction of Cardiovascular Disease using XGBoost with OPTUNA," *SN Comput Sci*, vol. 6, no. 5, p. 421, Apr. 2025, doi: 10.1007/s42979-025-03954-x.

[9] S. Dhanka and S. Maini, "A hybridization of XGBoost machine learning model by Optuna hyperparameter tuning suite for cardiovascular disease classification with significant effect of outliers and heterogeneous training datasets," *Int J Cardiol*, vol. 420, p. 132757, Feb. 2025, doi: 10.1016/j.ijcard.2024.132757.

[10] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," pp. 4765–4774, Nov. 2017.

[11] A. S. W. P. M. and D. R. Janosi, "Heart Disease," 1989, *UCI Machine Learning Repository*.