Jurnal Elektronika dan Teknik Informatika Terapan (JENTIK)
Vol. 4 No. 1 Maret 2026
e-ISSN : 2988-0874 p-ISSN : 2988-0866, Hal 48-58
DOI: https://doi.org/10.59061/jentik.v4i1.1517
Tersedia: https://e-journal.poltek-kampar.ac.id/index.php/JENTIK

Prediksi Risiko Penyakit Jantung Menggunakan *Xgboost* dengan Optimasi *Hyperparameter* Berbasis *GridsearchCV*

Syawaludin Khadafi Parinduri[1*], Antoni Pribadi[2], Nurkholis[3]
[1]Sistem Informasi, STIKOM Tunas Bangsa
[2]Teknik Informatika, Politeknik Kampar
[3]Teknologi Rekayasa Logistik, Politeknik Kampar
Email: syawaluddin@amiktunasbangsa.ac.id[1], antonipribadi.polkam@gmail.com[2], nurkholis11151996@gmail.com[3]

**Penulis Korespondensi: syawaluddin@amiktunasbangsa.ac.id*

Abstract. *Heart disease remains a leading cause of global morbidity and mortality, necessitating accurate and efficient risk prediction methods. This study aims to develop a robust heart disease prediction model using the XGBoost algorithm, focusing on performance enhancement through GridSearchCV -based hyperparameter optimization. The dataset was sourced from the UCI Machine Learning Repository, comprising 302 samples with 14 clinical and demographic patient attributes. The research methodology involved data preprocessing, stratified data splitting into training (75%) and testing (25%) sets, and implementation of the XGBoost model. GridSearchCV with 5-fold cross-validation was systematically applied to discover the optimal combination of hyperparameters. The optimization yielded the best configuration with learning_rate of 0.05, max_depth of 3, n_estimators of 200, min_child_weight of 5, gamma of 0.3, subsample of 0.8, and colsample_bytree of 1.0, achieving a mean cross-validation accuracy of 84.52%. Evaluation on the testing set showed an accuracy of 77.63%, precision of 78.57%, recall of 80.49%, F1-score of 79.52%, and ROC-AUC of 0.88. The high recall indicates the model's capability in detecting more than 80% of positive heart disease cases, which is critical in medical diagnosis contexts. The conclusion of this study is that the combination of XGBoost with a hyperparameter optimization strategy using GridSearchCV is an effective approach for heart disease risk prediction, potentially serving as a valuable diagnostic aid in clinical practice.*

Keywords: *GridSearchCV; Heart disease; Hyperparameter Optimization; Machine Learning; XGBoost.*

Abstrak. Penyakit jantung tetap menjadi penyebab utama morbiditas dan mortalitas global, mendorong kebutuhan akan metode prediksi risiko yang akurat dan efisien. Penelitian ini bertujuan untuk mengembangkan model prediksi penyakit jantung yang robust menggunakan algoritma XGBoost, dengan penekanan pada peningkatan kinerja melalui optimasi hyperparameter berbasis GridSearchCV. Data yang digunakan berasal dari UCI Machine Learning Repository, mencakup 302 sampel dengan 14 atribut klinis dan demografis pasien. Metode penelitian melibatkan prapemrosesan data, pembagian data secara stratified menjadi set pelatihan (training set, 75%) dan set pengujian (testing set, 25%), serta implementasi model XGBoost. GridSearchCV dengan validasi silang 5-fold diterapkan secara sistematis untuk mencari kombinasi hyperparameter optimal. Hasil optimasi menghasilkan konfigurasi terbaik dengan learning_rate sebesar 0,05, max_depth 3, n_estimators 200, min_child_weight 5, gamma 0,3, subsample 0,8, dan colsample_bytree 1,0, dengan rata-rata akurasi validasi silang (cross-validation) sebesar 84,52%. Evaluasi pada set pengujian menunjukkan akurasi sebesar 77,63%, presisi 78,57%, recall 80,49%, F1-score 79,52%, dan ROC-AUC 0,88. Recall yang tinggi menunjukkan kemampuan model dalam mendeteksi lebih dari 80% kasus positif penyakit jantung, yang kritis dalam konteks diagnosis medis. Kesimpulan dari studi ini adalah bahwa kombinasi XGBoost dengan strategi optimasi hyperparameter menggunakan GridSearchCV merupakan pendekatan yang efektif untuk prediksi risiko penyakit jantung, berpotensi menjadi alat bantu diagnostik yang berharga dalam praktik klinis.

Kata Kunci: GridSearchCV; Optimasi Hyperparameter; Pembelajaran Mesin; Penyakit Jantung; XGBoost.

1. PENDAHULUAN

Penyakit kardiovaskular menjadi penyebab kematian nomor satu di dunia dengan 17,9 juta jiwa pada tahun 2023 (Muharram et al., 2024). Di Indonesia, prevalensi hipertensi mencapai lebih dari 30% pada populasi dewasa (Harmadha et al., 2023; Arifin et al., 2022). Model risiko konvensional seperti Framingham Risk Score memiliki keterbatasan akurasi

Naskah Masuk: 27 Januari 2025; Revisi: 22 Februari 2026; Diterima: 28 Maret 2026; Terbit: 30 Maret 2026

untuk populasi dengan karakteristik genetik dan gaya hidup yang berbeda (Kurniawan et al., 2023). Kondisi ini mendorong pengembangan pendekatan prediktif berbasis machine learning (ML) yang mampu mengintegrasikan data klinis multidimensi secara objektif (Nia et al., 2023; Baghdadi et al., 2023).

Di antara berbagai algoritma ML, Extreme Gradient Boosting (XGBoost) menunjukkan kinerja unggul dalam prediksi penyakit jantung (Ali et al., 2023; Ivan and Prasetyo, 2023). XGBoost membangun pohon keputusan secara sekuensial dengan fungsi objektif yang mengintegrasikan regularisasi, sehingga mampu mengontrol kompleksitas model dan mencegah overfitting (Suhendra et al., 2023; Asih et al., 2023). Namun, kinerja XGBoost sangat bergantung pada konfigurasi hyperparameter internalnya (Anam et al., 2025; Dhanka and Maini, 2024). Grid Search Cross-Validation (GridSearchCV) merupakan metode optimasi yang sistematis dan ekshaustif untuk menemukan konfigurasi optimal (Muzayanah et al., 2024; Bernand et al., 2024).

Tinjauan literatur mengungkap beberapa celah: (1) optimasi simultan tujuh hyperparameter XGBoost menggunakan GridSearchCV masih terbatas (Ivan and Prasetyo, 2023; Bernand et al., 2024); (2) evaluasi komprehensif dengan berbagai metrik klinis masih terbatas (Bhatt et al., 2023; Asif et al., 2023); serta (3) studi pada populasi Indonesia masih relatif sedikit (Kurniawan et al., 2023; Septian et al., 2025).

Berdasarkan celah-celah tersebut, penelitian ini bertujuan untuk: (1) mengembangkan model prediksi risiko penyakit jantung menggunakan XGBoost; (2) menerapkan optimasi hyperparameter berbasis GridSearchCV; serta (3) mengevaluasi kinerja model menggunakan berbagai metrik. Kontribusi penelitian ini meliputi optimasi simultan tujuh hyperparameter XGBoost serta kontekstualisasi dalam kerangka populasi Indonesia.

2. KAJIAN TEORITIS

Penyakit Jantung dan *Machine Learning*

Penyakit jantung, khususnya penyakit arteri koroner, merupakan penyebab kematian utama secara global dengan 17,9 juta jiwa pada 2023 (Muharram et al., 2024). Di Indonesia, prevalensi hipertensi sebagai faktor risiko utama mencapai lebih dari 30% pada populasi dewasa (Harmadha et al., 2023). Diagnosis konvensional memiliki keterbatasan dalam mengidentifikasi risiko laten, mendorong eksplorasi pendekatan komputasional (Kurniawan et al., 2023).

Machine Learning (ML) menawarkan kemampuan menemukan pola kompleks dalam data klinis yang melampaui kapasitas analisis konvensional (Mahajan et al., 2023). Di antara berbagai pendekatan, *ensemble learning* secara konsisten menunjukkan kinerja unggul dalam prediksi kardiovaskular (Asif et al., 2023). Beberapa penelitian terdahulu yang relevan dirangkum pada Tabel 1.

Tabel 1. Ringkasan Penelitian Terdahulu

No.	Peneliti	Metode	Dataset	Hasil	Kelemahan
1	Ivan & Prasetyo (2023)	XGBoost + GridSearchCV	Heart Disease UCI	Akurasi 93,70%	Rentang hyperparameter yang digunakan masih terbatas sehingga belum mengeksplorasi kombinasi parameter secara optimal.
2	Asih et al. (2023)	XGBoost	Heart Disease UCI	Akurasi 92,50%	Tidak menerapkan optimasi hyperparameter sehingga performa model belum optimal.
3	Bernand et al. (2024)	XGBoost + Grid Search	Coronary Heart Disease Dataset	Akurasi 94,10%	Evaluasi hanya berfokus pada akurasi dan belum menggunakan metrik lain seperti Precision, Recall, F1-Score, dan AUC-ROC.
4	Mienye & Jere (2024)	Ensemble Learning + Bayesian Optimization	Heart Disease Dataset	Akurasi 95,20%	Penelitian berfokus pada metode ensemble sehingga tidak mengevaluasi performa XGBoost secara khusus.
5	Anam et al. (2025)	XGBoost + Bayesian Optimization (Optuna)	Health Risk Dataset	Akurasi 94,50%	Menggunakan Bayesian Optimization (Optuna) tanpa membandingkannya dengan metode GridSearchCV sehingga efektivitas masing-masing metode optimasi belum diketahui.

Berdasarkan tinjauan tersebut, teridentifikasi celah penelitian: (1) optimasi simultan tujuh *hyperparameter* XGBoost menggunakan GridSearchCV masih terbatas; (2) evaluasi komprehensif dengan berbagai metrik klinis masih terbatas; serta (3) kontekstualisasi dalam populasi Indonesia masih jarang (Kurniawan et al., 2023; Septian et al., 2025).

XGBoost dan *GridSearchCV*

XGBoost (*Extreme Gradient Boosting*) adalah algoritma *ensemble learning* berbasis *boosting* yang membangun pohon keputusan secara sekuensial, di mana setiap pohon baru memperbaiki kesalahan model sebelumnya (Ali et al., 2023). Fungsi objektif XGBoost mengintegrasikan fungsi kerugian dan regularisasi (Saputra et al., 2023):

dengan fungsi regularisasi dinyatakan sebagai:

$$L(\phi) = \sum_{\{i=1\}}^{n} l(y_i, \{y\}_i) + \sum_{\{k=1\}}^{k} \Omega(f_k)$$

di mana $\Omega(f) = \gamma T + \backslash frac\{1\}\{2\}\lambda \backslash sum_{\{j=1\}}^{\{T\}} w_j^2$ adalah istilah regularisasi yang mengontrol kompleksitas pohon (Ali et al., 2023). Keunggulan XGBoost meliputi efisiensi komputasi, penanganan *missing values* otomatis, serta regularisasi L1/L2 (Suhendra et al., 2023).

Grid Search Cross-Validation (GridSearchCV) mengeksplorasi secara ekshaustif semua

kombinasi *hyperparameter* dalam grid yang ditentukan (Bernand et al., 2024). Setiap kombinasi dievaluasi menggunakan *k-fold cross-validation* untuk estimasi kinerja yang robust (Muzayanah et al., 2024). Pendekatan ini memberikan reproduksibilitas tinggi dan cakupan menyeluruh terhadap ruang *hyperparameter* (Ivan and Prasetyo, 2023).

3. METODE PENELITIAN

Sumber Data

Penelitian ini menggunakan dataset *Heart Disease* dari UCI *Machine Learning Repository* melalui Kaggle (https://www.kaggle.com/datasets/johnsmith88/ heart-disease-dataset) (DeGroat et al., 2024). Dataset terdiri dari 302 sampel dengan 14 atribut (13 fitur dan 1 target biner), seluruhnya numerik. Tabel 2 menyajikan deskripsi seluruh atribut.

Tabel 2. Deskripsi Atribut Dataset Heart Disease

No	Atribut	Deskripsi	Tipe	Rentang
1	age	Usia pasien dalam tahun	Numerik	29–77
2	sex	Jenis kelamin (0 = perempuan, 1 = laki-laki)	Numerik	0–1
3	cp	Tipe nyeri dada (0–3)	Numerik	0–3
4	trestbps	Tekanan darah istirahat (mmHg)	Numerik	94–200
5	chol	Kadar kolesterol serum (mg/dl)	Numerik	126–564
6	fbs	Gula darah puasa > 120 mg/dl (0/tidak, 1/ya)	Numerik	0–1
7	restecg	Hasil EKG istirahat (0–2)	Numerik	0–2
8	thalach	Denyut jantung maksimum	Numerik	71–202
9	exang	Angina akibat olahraga (0/tidak, 1/ya)	Numerik	0–1
10	oldpeak	Depresi ST akibat olahraga	Numerik	0–6,2
11	slope	Kemiringan segmen ST (0–2)	Numerik	0–2
12	ca	Jumlah pembuluh darah utama (0–3)	Numerik	0–3
13	thal	Status thalassemia (0–3)	Numerik	0–3
14	target	Status penyakit jantung (0/tidak, 1/ada)	Numerik	0–1

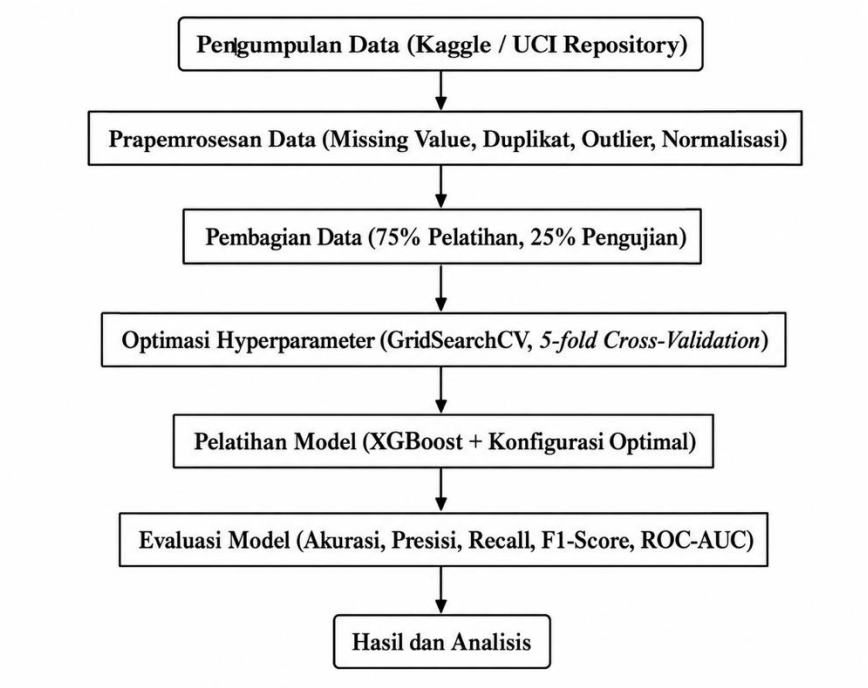
Tabel 3 menyajikan sampel 10 data dari dataset yang digunakan dalam penelitian ini.

Tabel 3. Sampel Data Heart Disease (10 Baris Pertama)

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
63	1	3	145	233	1	0	150	0	2,3	0	0	1	1
37	1	2	130	250	0	1	187	0	3,5	0	0	2	1
41	0	1	130	204	0	0	172	0	1,4	2	0	2	1
56	1	1	120	236	0	1	178	0	0,8	2	0	2	1
57	0	0	120	354	0	1	163	1	0,6	2	0	2	1
57	1	0	140	192	1	1	148	0	0,4	1	0	1	1
56	0	1	140	294	0	0	153	0	1,3	1	0	2	1
44	1	1	120	263	0	1	173	0	0,2	2	0	3	1
52	1	2	172	199	1	1	162	0	0,5	0	0	3	1
57	1	0	150	276	1	0	112	1	0,6	1	1	1	1

Alur Metodologi

Alur penelitian ini terdiri dari enam tahapan utama yang dilakukan secara berurutan dan sistematis, sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alur Metodologi Penelitian

Tahapan pertama adalah pengumpulan data dari repositori publik Kaggle yang bersumber dari UCI *Machine Learning Repository*. Tahapan kedua adalah prapemrosesan data yang meliputi identifikasi nilai hilang, deteksi duplikat, penanganan *outlier* menggunakan *Interquartile Range* (IQR), serta normalisasi fitur. Tahapan ketiga adalah pembagian data secara stratified menjadi set pelatihan (75%) dan set pengujian (25%). Tahapan keempat adalah optimasi *hyperparameter* menggunakan GridSearchCV dengan validasi silang 5-*fold* untuk menemukan konfigurasi optimal dari tujuh *hyperparameter* XGBoost. Tahapan kelima adalah pelatihan model XGBoost menggunakan konfigurasi optimal yang diperoleh dari tahapan sebelumnya. Tahapan terakhir adalah evaluasi kinerja model pada set pengujian menggunakan lima metrik: *accuracy*, *precision*, *recall*, *F1-score*, dan ROC-AUC (Ivan and Prasetyo, 2023; Asih et al., 2023).

Prapemrosesan Data

Prapemrosesan meliputi identifikasi nilai hilang, deteksi duplikat, serta penanganan *outlier* menggunakan *Interquartile Range* (IQR). Tidak ditemukan nilai hilang maupun duplikat pada dataset. Normalisasi diterapkan untuk memastikan konsistensi skala fitur (Asif et al., 2023; Asih et al., 2023).

Pembagian Data

Dataset dibagi secara stratified menjadi set pelatihan (75%, 226 sampel) dan set pengujian (25%, 76 sampel) (Mahajan et al., 2023). Pembagian stratified mempertahankan proporsi kelas target yang seimbang antar subset (Aryuni et al., 2023).

Optimasi Hyperparameter

Optimasi menggunakan GridSearchCV dengan validasi silang 5-fold mengeksplorasi secara ekshaustif semua kombinasi dari tujuh *hyperparameter* XGBoost (Bernand et al., 2024). Tabel 4 menyajikan deskripsi dan rentang nilai yang dieksplorasi.

Tabel 4. Hyperparameter XGBoost yang Dioptimasi

Hyperparameter	Deskripsi	Rentang Nilai
n_estimators	Jumlah pohon dalam ensemble	100, 200, 300, 500
learning_rate	Laju pembelajaran (kontribusi setiap pohon)	0,01; 0,05; 0,10; 0,20
max_depth	Kedalaman maksimum setiap pohon	3, 5, 7, 10
subsample	Proporsi sampel yang digunakan pada setiap pohon	0,6; 0,7; 0,8; 0,9; 1,0
colsample_bytree	Proporsi fitur yang digunakan pada setiap pohon	0,6; 0,7; 0,8; 0,9; 1,0
gamma	Ambang batas minimum pemisahan (*split*)	0; 0,1; 0,2; 0,3
min_child_weight	Bobot minimum pada simpul anak (*child node*)	1, 3, 5, 7

Evaluasi Kinerja Model

Evaluasi menggunakan lima metrik: *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan ROC-AUC (Asih et al., 2023). Penggunaan multi-metrik penting karena akurasi saja tidak cukup untuk data medis yang tidak seimbang (Asif et al., 2023). *Confusion matrix* dan kurva ROC dilengkapi untuk analisis detail pola kesalahan model.

4. HASIL DAN PEMBAHASAN

Hasil Prapemrosesan Data

Dataset *Heart Disease* dari UCI *Machine Learning Repository* terdiri dari 302 sampel dengan14 atribut (13 fitur dan 1 target). Tidak ditemukan nilai hilang maupun data duplikat, sehingga seluruh sampel dapat digunakan tanpa imputasi. Distribusi kelas target menunjukkan ketidakseimbangan, sehingga pembagian data dilakukan secara stratified dengan proporsi 75% pelatihan (226 sampel) dan 25% pengujian (76 sampel). Analisis korelasi menunjukkan bahwa *cp* (tipe nyeri dada), *thalach* (denyut jantung maksimum), *exang* (angina akibat olahraga), dan *oldpeak* (depresi ST) memiliki korelasi kuat dengan variabel target, konsisten dengan pemahaman medis mengenai faktor risiko kardiovaskular.

Hasil Optimasi Hyperparameter

GridSearchCV dengan validasi silang 5-fold ditunjukkan pada Tabel 5.

Tabel 5. Konfigurasi Hyperparameter Optimal Hasil GridSearchCV

Hyperparameter	Nilai Optimal
n_estimators	200
learning_rate	0,05
max_depth	3
min_child_weight	5
gamma	0,3
subsample	0,8
colsample_bytree	1,0

Best Accuracy Cross Validation: 84,52%. Konfigurasi ini menunjukkan bahwa model dengan *learning_rate* kecil (0,05) dan *max_depth* dangkal (3) menghasilkan generalisasi yang lebih baik. Regularisasi melalui *gamma* = 0,3 dan *min_child_weight* = 5 mencegah *overfitting*, sementara *subsample* = 0,8 meningkatkan keragaman ensemble.

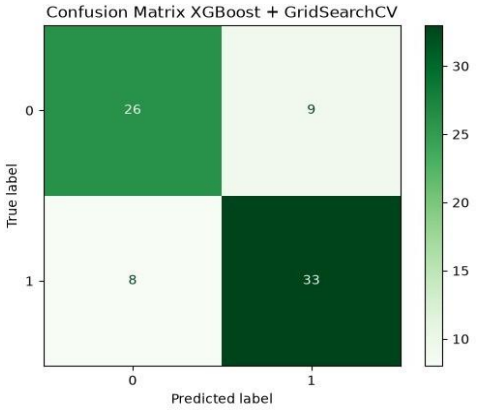
Hasil Evaluasi Kinerja Model

Evaluasi pada set pengujian menghasilkan metrik yang ditunjukkan pada Tabel 6.

Tabel 6. Hasil Evaluasi Kinerja Model XGBoost + GridSearchCV

Metrik	Nilai
Accuracy	77,63%
Precision	78,57%
Recall	80,49%
F1-Score	79,52%
ROC-AUC	0,88

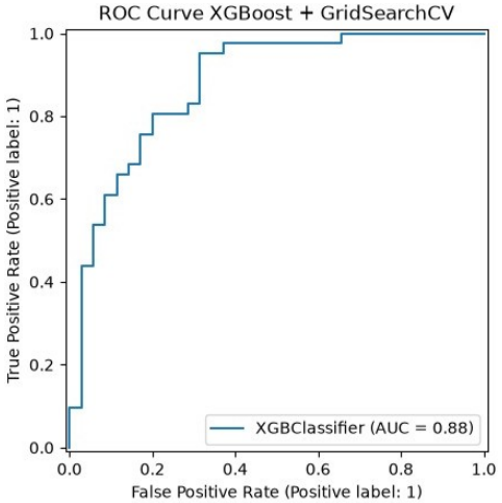
Confusion Matrix



Gambar 2. Confusion Matrix *XGBoost + GridSearchCV*

Confusion matrix menunjukkan TN = 26, FP = 9, FN = 8, dan TP = 33. False negative sebanyak 8 kasus merupakan kategori kritis dalam konteks klinis karena pasien yang tidak terdeteksi berisiko mengalami komplikasi serius.

ROC Curve



Gambar 3. ROC *Curve XGBoost + GridSearchCV*

AUC sebesar 0,88 menunjukkan probabilitas 88% bahwa model memberikan skor lebih tinggi pada pasien positif dibandingkan pasien negatif.

Pembahasan

Akurasi 77,63% dan ROC-AUC 0,88 menunjukkan kinerja yang memadai meskipun masih di bawah beberapa studi terdahulu seperti Ivan dan Prasetyo (2023) yang mencapai 93,70% dan Bernand et al. (2024) yang mencapai 94,10%. Perbedaan ini dapat dikaitkan dengan ukuran dataset yang relatif kecil (302 sampel) dan rentang *hyperparameter* yang berbeda. Namun, nilai ROC-AUC 0,88 sejalan dengan temuan studi lain yang menggunakan XGBoost dalam domain medis. Recall 80,49% merupakan temuan klinis yang signifikan karena *false negative* pada prediksi penyakit jantung memiliki konsekuensi jauh lebih serius dibandingkan *false positive*. Threshold klasifikasi dapat disesuaikan berdasarkan prioritas klinis: threshold lebih rendah meningkatkan sensitivitas untuk skrining awal.

Keunggulan GridSearchCV terletak pada sifatnya yang sistematis dan ekshaustif, memberikan reproduksibilitas yang lebih tinggi dibandingkan *Random Search*. Integrasi dengan 5-fold *cross-validation* memastikan konfigurasi yang dipilih robust terhadap variasi data. Keterbatasan utama meliputi ukuran dataset yang kecil (302 sampel), ketiadaan perbandingan langsung dengan model *baseline* lain, serta penggunaan dataset sekunder yang mungkin tidak sepenuhnya representatif terhadap populasi Indonesia. Penelitian selanjutnya disarankan untuk menggunakan dataset lebih besar, mengintegrasikan *Bayesian Optimization*, serta melakukan validasi eksternal pada populasi yang berbeda.

5. KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini telah mengembangkan model prediksi risiko penyakit jantung menggunakan algoritma Extreme Gradient Boosting (XGBoost) dengan optimasi hyperparameter berbasis Grid Search Cross-Validation (GridSearchCV). Berdasarkan hasil penelitian yang telah dilakukan, beberapa kesimpulan dapat ditarik sebagai berikut.

Pertama, optimasi hyperparameter menggunakan GridSearchCV dengan validasi silang 5-fold berhasil mengidentifikasi konfigurasi optimal yang terdiri dari n_estimators = 200, learning_rate = 0,05, max_depth = 3, min_child_weight = 5, gamma = 0,3, subsample = 0,8, dan colsample_bytree = 1,0, dengan rata-rata akurasi validasi silang sebesar 84,52%. Konfigurasi ini menunjukkan bahwa model dengan kompleksitas moderat dan regularisasi yang cukup kuat mampu menggeneralisasi dengan baik pada data yang belum pernah dilihat

sebelumnya.

Kedua, evaluasi model pada set pengujian menghasilkan akurasi sebesar 77,63%, presisi 78,57%, recall 80,49%, F1-score 79,52%, dan ROC-AUC 0,88. Recall yang tinggi (80,49%) menunjukkan kemampuan model dalam mendeteksi lebih dari 80% kasus positif penyakit jantung, yang merupakan aspek kritis dalam konteks diagnosis medis di mana false negative memiliki konsekuensi klinis yang serius. Nilai ROC-AUC 0,88 menunjukkan kemampuan diskriminatif yang baik dalam membedakan pasien dengan dan tanpa penyakit jantung.

Saran

Berdasarkan keterbatasan dan temuan penelitian ini, beberapa saran untuk penelitian selanjutnya adalah sebagai berikut. Pertama, penggunaan dataset berukuran lebih besar dari populasi Indonesia secara spesifik untuk meningkatkan representativitas dan generalisasi model. Kedua, eksplorasi teknik optimasi alternatif seperti Bayesian Optimization atau Optuna yang lebih efisien secara komputasi dibandingkan grid search ekshaustif. Ketiga, integrasi dengan teknik feature engineering dan seleksi fitur untuk meningkatkan akurasi prediksi. Keempat, validasi eksternal pada dataset klinis dari rumah sakit Indonesia untuk menguji ketertransferan model ke kondisi nyata. Kelima, pengembangan model ensemble yang menggabungkan XGBoost dengan algoritma lain untuk meningkatkan robustes prediksi.

DAFTAR PUSTAKA

Ali, Z. A., Abduljabbar, Z. H., Tahir, H. A., Sallow, A. B., and Almufti, S. M. (2023). extreme gradient boosting algorithm with machine learning: a review. Academic Journal of Nawroz University, 12(2).

Anam, S., Purwanto, I. N., Mahanani, D. M., Yusuf, F. I., and Rasikhun, H. (2025). Health risk classification using xgboost with bayesian hyperparameter optimization. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi).

Arifin, H., Chou, K., Ibrahim, K., Fitri, S. U. R., Pradipta, R. O., Rias, Y. A., Sitorus, N., Wiratama, B. S., Setiawan, A., Setyowati, S., Kuswanto, H., Mediarti, D., Rosnani, R., Sulistini, R., and Pahria, T. (2022). Analysis of modifiable, non-modifiable, and physiological risk factors of non-communicable diseases in indonesia: Evidence from the 2018 indonesian basic health research. Journal of Multidisciplinary Healthcare.

Aryuni, M., Adiarto, S., Miranda, E., Madyatmadja, E. D., Albert, V. D. S., and Sestomi, E. (2023). Imbalanced learning in heart disease categorization: Improving minority class prediction accuracy using the smote algorithm. INTERNATIONAL JOURNAL of FUZZY LOGIC and INTELLIGENT SYSTEMS, 23(2):140–151.

Asif, D., Bibi, M., Arif, M. S., and Mukheimer, A. (2023). Enhancing heart disease prediction through ensemble learning techniques with hyperparameter optimization. Algorithms, 16(6):308.

Asih, P. S., Azhar, Y., Wicaksono, G. W., and Akbi, D. R. (2023). Interpretable machine

learning model for heart disease prediction. Procedia Computer Science.

Baghdadi, N. A., Abdelaliem, S. M. F., Malki, A., Gad, I., Ewis, A. A., and Atlam, E.-S. (2023). Advanced machine learning techniques for cardiovascular disease early detection and diagnosis. Journal of Big Data, 10:101.

Bernand, C., Mirand, E., and Aryun, M. (2024). Coronary heart disease prediction models using machine learning and deep learning algorithms. In PROCEEDINGS OF THE 8TH INTERNATIONAL CONFERENCE ON ENGINEERING, TECHNOLOGY, AND INDUSTRIAL APPLICATIONS 2021 (8th ICETIA 2021): Engineering, Environment, and Health: Exploring the Opportunities for the Future.

Bhatt, C., Patel, P. V., Ghetia, T., and Mazzeo, P. L. (2023). Effective heart disease prediction using machine learning techniques. Algorithms, 16(2):88.

DeGroat, W., Abdelhalim, H., Patel, K., Mendhe, D., Zeeshan, S., and Ahmed, Z. (2024). Discovering biomarkers associated and predicting cardiovascular disease with high accuracy using a novel nexus of machine learning techniques for precision medicine. Scientific Reports, 14:1555.

Dhanka, S. and Maini, S. (2024). 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. International Journal of Cardiology, page 132757.

Harmadha, W. S. P., Muharram, F. R., Gaspar, R. S., Azimuth, Z., Sulistya, H. A., Firmansyah, F. S., Multazam, C. E. C. Z., Harits, M., and Putra, R. M. (2023). Explaining the increase of incidence and mortality from cardiovascular disease in indonesia: A global burden of disease study analysis (2000–2019). PLoS ONE.

Ivan, J. and Prasetyo, S. Y. (2023). Heart disease prediction using ensemble model and hyperparameter optimization. International Journal on Recent and Innovation Trends in Computing and Communication.

Kurniawan, R., Utomo, B., Siregar, K. N., Ramli, K., Besral, B., Suhatril, R. J., and Pratiwi, O. A. (2023). Hypertension prediction using machine learning algorithm among indonesian adults. IAES International Journal of Artificial Intelligence (IJ-AI), 12(2):776.

Mahajan, P., Uddin, S., Hajati, F., and Moni, M. A. (2023). Ensemble learning for disease prediction: A review. Healthcare, 11(12):1808.

Mienye, I. D. and Jere, N. (2024). Optimized ensemble learning approach with explainable ai for improved heart disease prediction. Information, 15(7):394.

Muharram, F. R., Multazam, C. E. C. Z., Mustofa, A., Socha, W., Andrianto, A., Martini, S., Aminde, L. N., and Yi-Li, C. (2024). The 30 years of shifting in the indonesian cardiovascular burden—analysis of the global burden of disease study. Journal of Epidemiology and Global Health.

Muzayanah, R., Pertiwi, D. A. A., Ali, M., and Muslim, M. A. (2024). Comparison of gridsearchcv and bayesian hyperparameter optimization in random forest algorithm for diabetes prediction. Journal of Soft Computing Exploration.

Nia, N. G., Kaplanoglu, E., and Nasab, A. (2023). Evaluation of artificial intelligence techniques in disease diagnosis and prediction. Discover Artificial Intelligence, 3:10.

Saputra, J., Lawrencya, C., Saini, J. M., and Suharjito, S. (2023). Hyperparameter optimization for cardiovascular disease data-driven prognostic system. Visual Computing for Industry, Biomedicine, and Art.

Septian, E., Khaefi, M. R., Athoillah, A., Aisyah, D. N., Hardhantyo, M., Rahman, F. M., and Manikam, L. (2025). Prediction of personalised hypertension using machine learning in indonesian population. Journal of Medical Systems, 49(1).

Suhendra, R., Husdayanti, N., Suryadi, S., Juliwardi, I., Sanusi, S., Ridho, A., Ardiansyah, M., Murhaban, M., and Ikhsan, I. (2023). Cardiovascular disease prediction using gradient boosting classifier. Infolitika Journal of Data Science.