JNATIA Volume 2, Nomor 3, Mei 2024 p-ISSN: 2986-3929
Jurnal Nasional Teknologi Informasi dan Aplikasinya e-ISSN: 3032-1948

Klasifikasi Penyakit Gagal Jantung dengan Algoritma XGBoost

I Gusti Gde Bagus Bhadrika Artawibawa[a1], AAIN Eka Karyawati[a2]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
[1]artawibawa.2208561052@student.unud.ac.id
[2]eka.karyawati@unud.ac.id

Abstract

Heart, as one of the most important organs in the body, carries a risk of death if abnormalities occur. Heart problems are divided into two categories: heart failure and heart attacks. According to WHO data, approximately 7.3 million people worldwide die due to heart disease. This study uses a dataset of heart disease patients and applies the XGBoost algorithm. The objectives of this study are to process and analyze the data, implement the XGBoost algorithm for heart disease classification, and evaluate the performance of the XGBoost algorithm. The result of this study is the performance evaluation of the XGBoost algorithm, which achieved an accuracy of 93%.

Keywords: *Heart, Classification, Failure, XGBoost, Disease*

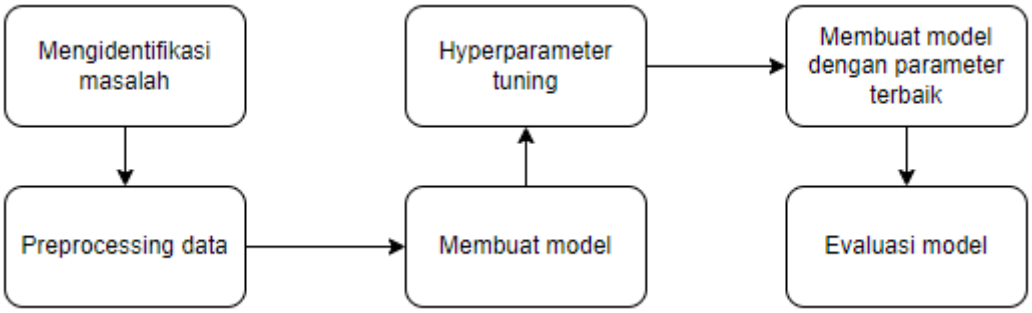
1. Pendahuluan

Penyakit kardiovaskular merupakan penyakit pada jantung dan pembuluh darah sebagai penyebab terjadinya gagal jantung. Penyakit ini sering terjadi dan merupakan salah satu penyebab utama kematian di dunia. Berdasarkan data dari WHO, pada tahun 2021 angka kematian akibat penyakit jantung mencapai 17,8 juta orang atau satu dari tiga kematian di dunia pada tiap tahunnya disebabkan oleh penyakit jantung. Gagal jantung adalah kondisi jantung kehilangan kemampuannya untuk memompa darah untuk jumlah yang cukup dalam memenuhi kebutuhan metabolisme tubuh atau jantung hanya mampu melakukannya dengan tekanan pengisian yang tinggi atau dapat juga terjadi kedua duanya secara bersamaan. Beberapa faktor yang biasanya menyebabkan penyakit gagal jantung diantaranya adalah diabetes, tekanan darah tinggi pola hidup yang tidak sehat dan kurangnya aktivitas fisik [1]. Oleh sebab itu perlu adanya suatu model yang akurat untuk mengklasifikasikan gagal jantung berdasarkan informasi klinis dan gaya hidup pasien pengidap penyakit tersebut, sebagai solusi alternatif dalam pemberian obat yang tepat. Akan tetapi, tingkat akurasi klasifikasi kejadian terkait gagal jantung dalam praktik klinis biasanya kurang sensitif. Saat ini, banyak cara untuk melakukan prediksi, salah satunya dengan membuat pemodelan menggunakan algoritma XGBoost, yaitu pengembangan algoritma dan model secara statistik yang menggunakan sistem komputer serta membutuhkan data yang mengandalkan pola serta inferensi [1]. Dalam beberapa penelitian yang memiliki kemiripan dalam melakukan klasifikasi penyakit jantung, metode-metode yang digunakan umumnya menghasilkan nilai akurasi yang berbeda. Penelitian Perancangan Sistem Klasifikasi Penyakit Jantung Mengunakan Naïve Bayes yang diterbitkan pada tahun 2019[6]. Menggunakan metode Naive Bayes menghasilkan nilai rata rata akurasi sebesar 90,61%. Penelitian Analisis Penyakit Jantung Koroner Menggunakan Decision Tree yang diterbitkan pada tahun 2019 menjelaskan bahwa dengan menggunakan metode CART Decision Tree, mendapatkan akurasi sebesar 80% [7]. Penelitian Prediction of Heart Diseases using Random Forest yang diterbitkan Maret 2021 menjelaskan bahwa penerapan klasifikasi pada 303 data penyakit jantung menggunakan algoritma Random Forest, mendapatkan hasil akurasi sebesar 86.69% [2]. Tujuan dari penelitian ini untuk mengidentifikasi pasien gagal jantung berdasarkan informasi klinis dan gaya hidup pasien dengan algoritma XGBoost dan memberikan sebuah algoritma atau model yang dapat digunakan untuk mengklasifikasi penyakit gagal jantung.

p-ISSN: 2986-3929
e-ISSN: 3032-1948

2. Metode Penelitian

Penelitian dibagi menjadi beberapa tahapan yaitu yang pertama ada mengidentifikasi masalah, preprocessing data, membuat model, melakukan hyperparameter tuning, membuat model dengan parameter terbaik, dan evaluasi model. Untuk gambaran dari tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari Kaggle, dan untuk datasetnya dapat dilihat pada Gambar 2. https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction/data

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
0	40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
1	49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
2	37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
3	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
4	54	M	NAP	150	195	0	Normal	122	N	0.0	Up	0
...	...	...	...	...	...	...	...	...	...	...	...	...
913	45	M	TA	110	264	0	Normal	132	N	1.2	Flat	1
914	68	M	ASY	144	193	1	Normal	141	N	3.4	Flat	1
915	57	M	ASY	130	131	0	Normal	115	Y	1.2	Flat	1
916	57	F	ATA	130	236	0	LVH	174	N	0.0	Flat	1
917	38	M	NAP	138	175	0	Normal	173	N	0.0	Up	0

918 rows × 12 columns

Gambar 2. Dataset yang digunakan

.
Data yang digunakan berjumlah 918 baris dan 12 kolom, didalamnya terdapat beberapa parameter seperti *age, sex, cholesterol, chestpaintype,* dan *oldpeak*.

2.2. Preprocessing Data

Sebelum data digunakan untuk melatih model atau algoritma XGBoost, diperlukan adanya preprocessing data agar tidak berdampak buruk pada performa dari model tersebut. Pada penelitian ini, preprocessing data mencakup pengecekan terhadap data duplikat, dan melakukan label encoder untuk data yang bersifat kategorikal. Setelah dilakukan preprocessing data, kemudian data tersebut akan dibagi menjadi data training sebesar 80% dan data testing sebesar 20%.

2.3. Algoritma XGBoost

Algoritma Extreme Gradient Boosting (XGBoost) merupakah salah satu metode yang unik dimana XGBoost menggabungkan proses boostingdengan gradient boosting untuk mendapatkan hasil 10 kali lebih cepat. Dimana Extreme Gradient Boosting menggabungkan tiga macam ensembles algorithm yaitu classification dan regression trees (CART) [3]. XGBoost Sebenarnya adalah algoritma yang lebih banyak digunakan dalam pemodelan prediktif, seperti regresi dan klasifikasi. Meskipun XGBoost dapat menghasilkan skor kepentingan atribut, biasanya digunakan untuk menentukan atribut yang paling penting dalam pemodelan prediktif, bukan untuk seleksi atribut untuk clustering K-Means. Meskipun demikian, XGBoost dapat digunakan sebagai alat tambahan untuk analisis atribut pada clustering K-Means. Berikut tahapan penerapannya:

a. Melakukan clustering menggunakan algoritma K-Means untuk membagi data menjadi klaster-klaster tertentu.
b. Menggabungkan hasil clustering dengan dataset asli
c. Memisahkan atribut dan label klaster
d. Membangun model XGBoost untuk memperbaiki clustering
e. Menghitung skor kepentingan setiap atribut
f. Menampilkan atribut yang paling penting berdasarkan skor kepentingan tertinggi [5].

2.4. Hyperparameter Tuning

Hyperparameter adalah parameter yang ditetapkan sebelum proses training dimulai. Parameter ini dapat disesuaikan dan secara langsung dapat mempengaruhi seberapa baik kinerja model. Untuk menemukan hyperparameter yang paling optimal, berbagai strategi diperlukan untuk menyesuaikannya. Cara yang paling mudah adalah dengan mencoba berbagai kombinasi hyperparameter. Namun, seiring waktu, banyak pendekatan telah diusulkan untuk mengoptimalkan hyperparameter ini, seperti Grid Search dan Random Search [4]. Pada penelitian ini proses tuning hyperparameter dilakukan menggunakan metode Randomized Search Cross-Validation (Randomized Search CV), yang dipilih karena efisiensinya dalam mengeksplorasi ruang hyperparameter yang besar tanpa memerlukan pencarian lengkap seperti pada Grid Search.

3. Hasil dan Diskusi

Pada penelitian ini digunakan algoritma XGBoost dengan hyper paramerter tuning. Dataset yang digunakan dibagi menjadi 80% training dan 20% testing. Sebelum melakukan data training, dilakukan pengecekan terhadap data duplikat, dan melakukan label encoder untuk data yang bersifat kategorikal. Pembuatan model akan didasarkan pada parameter terbaik. Selanjutnya penggunaan K-fold cross validation yang bertujuan untuk mendapatkan hasil penilaian akurasi yang lebih maksimal.

3.1. Performa Algoritma XGBoost dengan Hyper Parameter Tuning

Berikut merupakan hasil dari algoritma XGBoost. dapat dilihat pada gambar 4 hasil metrics score dari algoritma ini menghasilkan akurasi k-fold cross validation menghasilkan sebesar 88% akurasi dari model XGBoost sebesar 83%, nilai precision sebesar 93%, Recall sebesar 93%, dan F1 score sebesar 93%. Hasil dari penilaian metrics score dari algoritma ini tergolong baik untuk proses klasifikasi. Untuk lebih lengkapnya dapat dilihat pada Gambar 3 dibawah ini.

p-ISSN: 2986-3929
e-ISSN: 3032-1948

	precision	recall	f1-score	support
0	0.92	0.93	0.92	86
1	0.94	0.93	0.93	98
accuracy			0.93	184
macro avg	0.93	0.93	0.93	184
weighted avg	0.93	0.93	0.93	184

Gambar 3. Hasil metrics score

Pada Gambar 4 menunjukkan confusion matrix dari model klasifikasi XGBoost dengan hyperparameter tuning RandomizedSearchCV
.

Gambar 4. Confusion Matrix XGBoost + RandomizedSearchCV

4. Kesimpulan

Berdasarkan hasil penelitian, ditemukan bahwa algoritma XGBoost dapat digunakan untuk mengklasifikasikan penyakit gagal jantung. Performa dari algoritma ini dengan dilakukannya hyperparameter tuning mendapatkan akurasi sebesar 93%, precision sebesar 93%, recall sebesar 93%, dan f1-score sebesar 93%. Jadi, kesimpulannya jika dibandingkan dengan algoritma Decision Tree yang hanya mendapatkan akurasi sebesar 80%, penerapan XGBoost dengan hyperparameter tuning melalui RandomizedSearchCV adalah salah satu pendekatan yang sangat efektif untuk klasifikasi penyakit gagal jantung dengan akurasi mencapai 93%. Hasil penelitian ini menunjukkan potensi besar dalam membantu tenaga medis untuk mendeteksi dan mendiagnosis gagal jantung dengan akurasi yang tinggi, yang pada akhirnya dapat berkontribusi pada peningkatan kualitas perawatan pasien.

Daftar Pustaka

[1] R. Arisandi, “Perbandingan Model Klasifikasi Random Forest Dengan Resampling Dan Tanpa Resampling Pada Pasien Penderita Gagal Jantung,” *Jurnal Gaussian: Jurnal Statistika Undip*, vol. 12, no. 1, pp. 136–145, May 2023, doi: 10.14710/j.gauss.12.1.136-145.

[2] D. H. Depari, Y. Widiastiwi, and M. M. Santoni, “Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung,” Informatik, vol. 18, no. 3, p. 239, Dec. 2022, doi: 10.52958/iftk. v18i3.4694.

[3] Y. F. Wijaya and S. Y. J. Prasetyo, “Model penilaian tata guna lahan dengan citra Landsat 8 OLI menggunakan algoritma XGBOOST diwilayah beresiko tsunami (Studi kasus: Kota Palu Sulawesi Tengah),” ICM (Indonesian Journal of Computing and Modeling), vol. 4, no. 1, pp. 23–28, Jul. 2021, doi: 10.24246/icm. v4i1.4981.

[4] N. F. Rahmadayana and N. Y. Sibaroni, “Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization,” Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), vol. 5, no. 5, pp. 936–942, Oct. 2021, doi: 10.29207/resti. v5i5.3457.
[5] Bengnga, Amiruddin, and Rezqiwati Ishak. "Penerapan XGBoost untuk Seleksi Atribut pada K-Means dalam Clustering Penerima KIP Kuliah." *Jambura Journal of Electrical and Electronics Engineering* 5, no. 2 (2023): 192-196.
[6] M. A. Bianto, K. Kusrini, and S. Sudarmawan, “Perancangan Sistem klasifikasi penyakit jantung mengunakan Naïve Bayes,” *Citec (Creative Information Technology) Journal/Citec Journal*, vol. 6, no. 1, p. 75, Apr. 2019, doi: 10.24076/citec.2019v6i1.231.
[7] Akbar, Jimmi Afriando, Zidane Ibrahim Fadela, Luthfi Fachruddin, Feby Ardiansyah, and Zuhdi Mukarram Bakhri. "Analisis Penyakit Jantung Koroner Menggunakan Decision Tree." *Departemen Ilmu Komputer, IPB University, Cvd* (2019): 3-7.

p-ISSN: 2986-3929
e-ISSN: 3032-1948

Halaman ini sengaja dibiarkan kosong