

**IMPLEMENTASI METODE EXTREME GRADIENT BOOSTING
(XGBOOST) UNTUK KLASIFIKASI PENYAKIT JANTUNG**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

DAMA DAMANIK

20.11.3642

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**IMPLEMENTASI METODE EXTREME GRADIENT BOOSTING
(XGBOOST) UNTUK KLASIFIKASI PENYAKIT JANTUNG**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

DAMA DAMANIK

20.11.3642

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024**

HALAMAN PERSETUJUAN

SKRIPSI

**IMPLEMENTASI METODE EXTREME GRADIENT BOOSTING (XGBOOST)
UNTUK KLASIFIKASI PENYAKIT JANTUNG**

yang disusun dan diajukan oleh

Dama Damanik

20.11.3642

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 23 Januari 2024

Dosen Pembimbing,



Ferian Fauzi Abdulloh, M.Kom.

NIK. 190302276

HALAMAN PENGESAHAN
SKRIPSI
IMPLEMENTASI METODE EXTREME GRADIENT BOOSTING
(XGBOOST) UNTUK KLASIFIKASI PENYAKIT JANTUNG

yang disusun dan diajukan oleh

Dama Damanik

20.11.3642

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Januari 2024

Susunan Dewan Penguji

Nama Penguji

Donni Prabowo, M.Kom
NIK. 190302253

Rizqi Sukma Kharisma, M.Kom
NIK. 190302215

Ferian Fauzi Abdulloh, M.Kom
NIK. 190302276

Tanda-Tangan

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 Januari 2024

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Dama Damanik
NIM : 20.11.3642

Menyatakan bahwa Skripsi dengan judul berikut:

IMPLEMENTASI METODE EXTREME GRADIENT BOOSTING (XGBOOST) UNTUK KLASIFIKASI PENYAKIT JANTUNG

Dosen Pembimbing : Ferian Fauzi Abdulloh, M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 23 Januari 2024

Yang Menyatakan,



Dama Damanik

HALAMAN PERSEMBAHAN

Alhamdulillah rabbil 'alamin, puji syukur atas kehadiran Allah subhanahu wa ta'ala yang telah memberikan rahmat, karunia dan hidayah-Nya sehingga saya dapat menyelesaikan tugas skripsi ini dengan sebaik-baiknya.

Dengan penuh rasa syukur, skripsi ini saya persembahkan kepada Allah subhanahu wa ta'ala yang telah memberikan saya segala kenikmatan sehingga saya memiliki kemampuan untuk menyelesaikan skripsi ini.

Dan selanjutnya, skripsi ini saya persembahkan untuk keluarga saya, yaitu Bapak, Mama dan Kakak-Kakak saya atas doa, pengertian dan dukungan yang telah diberikan sehingga saya dapat menyelesaikan studi dan menjadi lulusan sarjana.

Dan tidak lupa juga kepada semua pihak yang telah membantu saya dalam penyusunan skripsi ini yang tidak dapat disebutkan satu persatu, semoga Allah subhanahu wa ta'ala meridhoi kita semua.

KATA PENGANTAR

Alhamdulillah rabbil 'alamin, segala puji dan syukur penulis panjatkan atas kehadiran Allah subhanahu wa ta'ala atas segala rahmat dan karunia-Nya sehingga penulis mampu menyelesaikan skripsi dengan judul “IMPLEMENTASI METODE EXTREME GRADIENT BOOSTING (XGBOOST) UNTUK KLASIFIKASI PENYAKIT JANTUNG”. Serta tidak lupa juga shalawat beserta salam kepada baginda Nabi Muhammad shallallahu 'alaihi wasallam beserta keluarganya, para sahabatnya dan ahli sunnah.

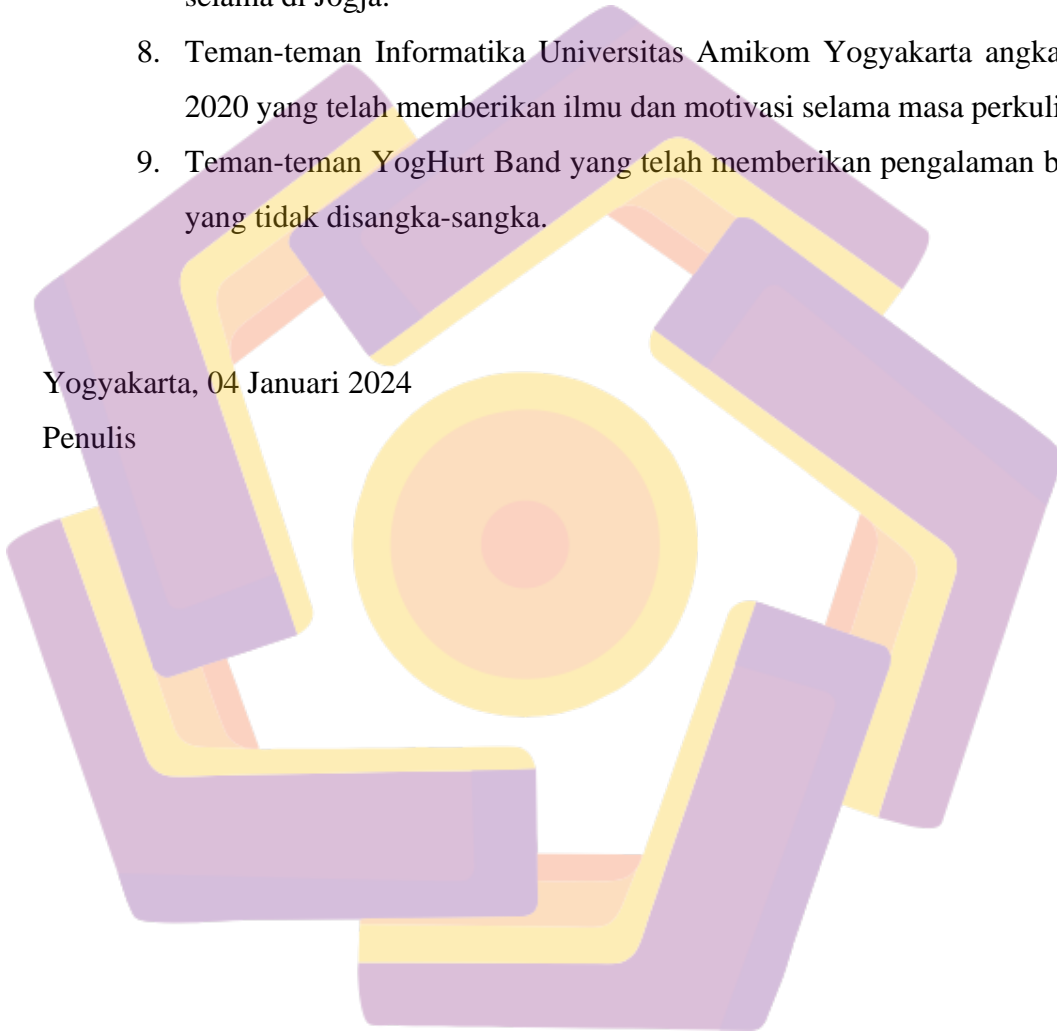
Skripsi ini ditulis untuk memenuhi persyaratan untuk mencapai gelar Sarjana Komputer dalam Program Studi Informatika Fakultas Ilmu Komputer. Selama penulisan skripsi ini, penulis telah banyak mendapatkan bimbingan dan bantuan dari berbagai pihak, baik berupa kritik, saran, motivasi serta bantuan lainnya. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Kedua orang tua dan kakak-kakak saya yang telah memberikan dukungan, doa, motivasi dan pendidikan sampai ke jenjang perguruan tinggi ini.
2. Bapak Prof. Dr. M. Suyanto, M.M, selaku Rektor Universitas AMIKOM Yogyakarta.
3. Bapak Hanif Al Fatta, M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta beserta seluruh jajarannya.
4. Ibu Windha Mega Pradnya Duhita, M.Kom, selaku Ketua Program Studi Informatika Universitas AMIKOM Yogyakarta beserta seluruh jajarannya.
5. Bapak Ferian Fauzi Abdullah, M.Kom, selaku dosen pembimbing saya yang telah memberikan waktu, saran, kritik, motivasi dan bimbingan dalam skripsi ini sehingga skripsi ini dapat selesai dengan baik.

6. Bapak Ibu Dosen Program Studi Informatika Universitas AMIKOM Yogyakarta yang telah mengajari saya dari semester awal hingga akhir sehingga saya memperoleh banyak ilmu dan pengalaman.
7. Saudara-saudara saya : Hijra, Midun, Rafi, Ardhi, Faris, Rama, Dendy, Muiz, Rizky dan Mugen yang selalu menemani dan berbagi suka duka selama di Jogja.
8. Teman-teman Informatika Universitas Amikom Yogyakarta angkatan 2020 yang telah memberikan ilmu dan motivasi selama masa perkuliaan.
9. Teman-teman YogHurt Band yang telah memberikan pengalaman baru yang tidak disangka-sangka.

Yogyakarta, 04 Januari 2024

Penulis



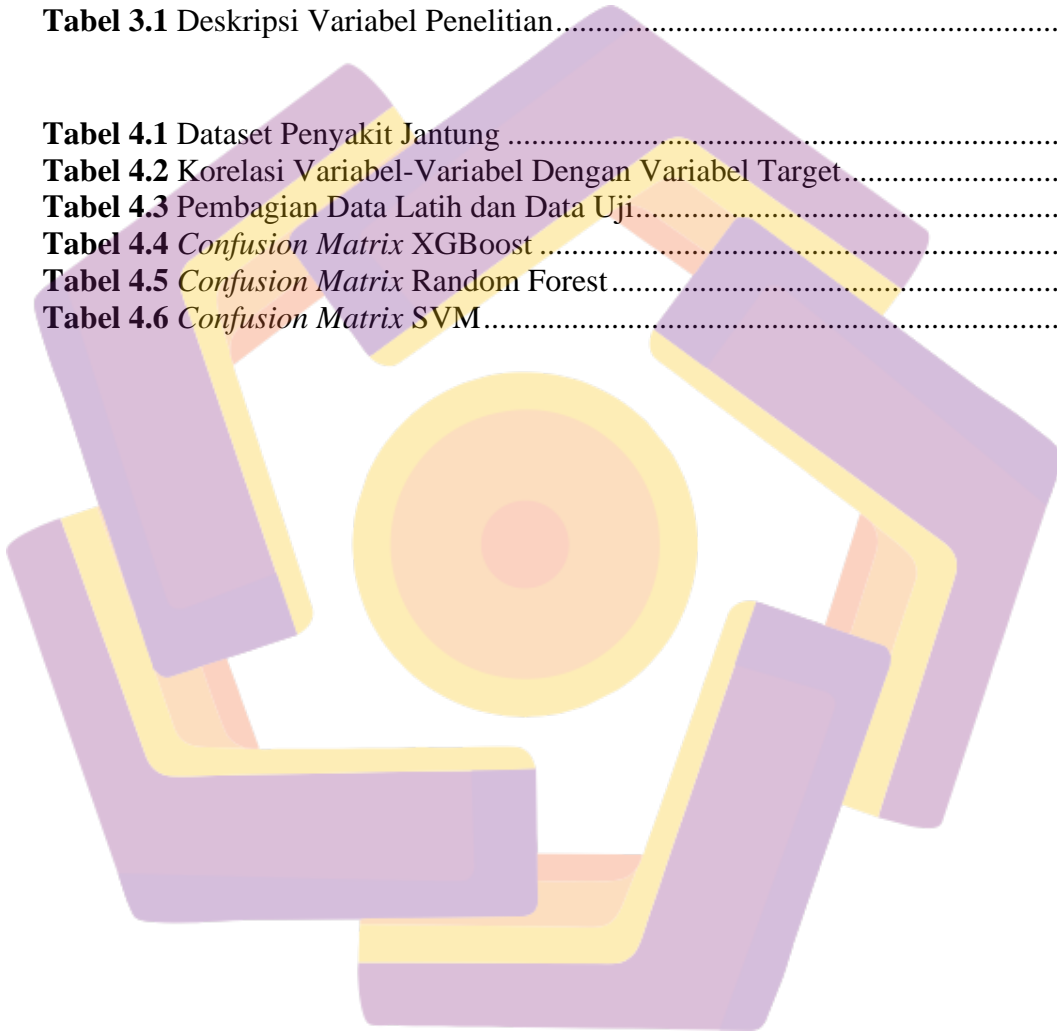
DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR ISTILAH	xii
INTISARI	xiii
<i>ABSTRACT</i>	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	15
2.2.1. Penyakit Jantung	15
2.2.2. Exploratory Data Analyst	15
2.2.3. Feature Selection.....	16
2.2.4. Pearson Correlation.....	17
2.2.5. Machine Learning	17

2.2.6.	Klasifikasi	18
2.2.7.	Extreme Gradient Boosting.....	19
2.2.8.	Random Forest	20
2.2.9.	SVM.....	21
2.2.10.	Confusion Matrix	22
BAB III METODE PENELITIAN		24
3.1	Alur Penelitian	24
3.2	Pengumpulan Data.....	26
3.3	Alat dan Bahan.....	29
BAB IV HASIL DAN PEMBAHASAN		30
4.1	Exploratory Data Analayst (EDA).....	30
4.2	Filtering Data	35
4.3	Pembagian Data Latih dan Data Uji	37
4.4	Klasifikasi	38
4.4.1.	Xtreme Gradient Boosting	38
4.4.2.	Random Forest	38
4.4.3.	SVM.....	39
4.5	Evaluasi.....	39
4.5.1.	XGBoost	39
4.5.2.	Random Forest	41
4.5.3.	SVM.....	42
4.6	Perbandingan Model XGBoost, Random Forest dan SVM.....	44
BAB V PENUTUP		45
5.1	Kesimpulan	45
5.2	Saran	45
REFERENSI		46

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian.....	10
Tabel 2.2 Parameter Pada XGBoost.....	20
Tabel 2.3 Ilustrasi Confusion Matrix	23
Tabel 3.1 Deskripsi Variabel Penelitian.....	26
Tabel 4.1 Dataset Penyakit Jantung	30
Tabel 4.2 Korelasi Variabel-Variabel Dengan Variabel Target.....	36
Tabel 4.3 Pembagian Data Latih dan Data Uji.....	37
Tabel 4.4 <i>Confusion Matrix</i> XGBoost	39
Tabel 4.5 <i>Confusion Matrix</i> Random Forest.....	41
Tabel 4.6 <i>Confusion Matrix</i> SVM.....	42

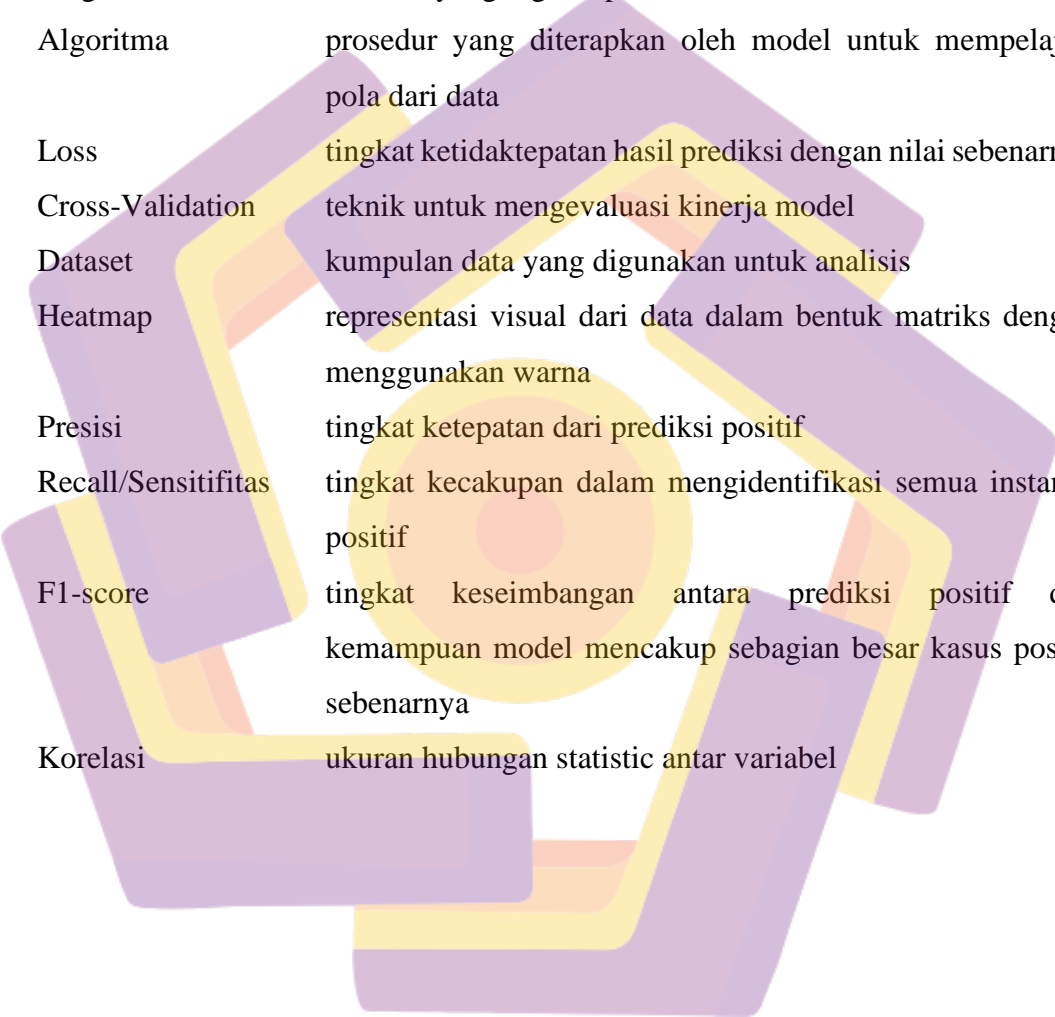


DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian.....	24
Gambar 3.2 Variabel-Variabel Penelitian	28
Gambar 4.1 Tampilan Dataset.....	30
Gambar 4.2 Barplot Usia Sampel.....	31
Gambar 4.3 Barplot Usia Sampel Dengan Penyakit Jantung.....	32
Gambar 4.4 Barplot Usia Sampel Tidak Terkena Penyakit Jantung	32
Gambar 4.5 Pieplot Jenis Nyeri Dada	33
Gambar 4.6 Barplot Tekanan Darah Istirahat	34
Gambar 4.7 Barplot Kolesterol	34
Gambar 4.8 Heatmap Feature Selection Dataset.....	36
Gambar 4.9 Perbandingan Akurasi dan F1-score Metode XBoost, Random Forest dan SVM.....	44



DAFTAR ISTILAH



Akurasi	tingkat ketepatan prediksi model dengan nilai sebenarnya
Data Training	sampel yang digunakan untuk melatih model
Data Testing	sampel yang digunakan untuk menguji model
Target	variabel yang ingin diprediksi oleh model
Algoritma	prosedur yang diterapkan oleh model untuk mempelajari pola dari data
Loss	tingkat ketidaktepatan hasil prediksi dengan nilai sebenarnya
Cross-Validation	teknik untuk mengevaluasi kinerja model
Dataset	kumpulan data yang digunakan untuk analisis
Heatmap	representasi visual dari data dalam bentuk matriks dengan menggunakan warna
Presisi	tingkat ketepatan dari prediksi positif
Recall/Sensitifitas	tingkat kecakupan dalam mengidentifikasi semua instance positif
F1-score	tingkat keseimbangan antara prediksi positif dan kemampuan model mencakup sebagian besar kasus positif sebenarnya
Korelasi	ukuran hubungan statistic antar variabel

INTISARI

Penyakit jantung merupakan tantangan kesehatan serius yang memerlukan prediksi dan diagnosis yang akurat. Penelitian ini mengeksplorasi penerapan metode eXtreme Gradient Boosting (XGBoost) dalam klasifikasi dan prediksi penyakit jantung dengan menggunakan dataset klinis pasien yang didapatkan dari Kaggle (<https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>). Fokus penelitian melibatkan analisis dampak implementasi XGBoost, tingkat akurasi yang dicapai, identifikasi variabel yang signifikan, dan pemilihan parameter optimal. Melibatkan perbandingan dengan algoritma Random Forest dan Support Vector Machine (SVM). *Exploratory Data Analyst* menggambarkan distribusi karakteristik seperti usia, jenis nyeri dada, tekanan darah, dan kolesterol pada sampel. *Feature selection* menyoroti variabel yang paling mempengaruhi output, termasuk *exng*, *cp*, *oldpeak*, dan *thalachh*. Hasil evaluasi kinerja model XGBoost menggunakan Confusion Matrix menunjukkan akurasi yang baik (85,25%), presisi yang memadai (81,57%), recall yang tinggi (93,94%), dan f1-score sebesar 87,32%. Dalam perbandingannya SVM menghasilkan nilai yang paling baik dengan akurasi 86,67% dan f1-score 87,50% mengungguli performa XGBoost dengan akurasi 85,25% dan f1-score 87,32% dan Random Forest dengan akurasi 81,97% dan f1-score 85,34%. Kesimpulan penelitian menekankan bahwa XGBoost efektif dalam klasifikasi penyakit jantung, dan saran untuk penelitian selanjutnya mencakup implementasi metode pada dataset lain dan penanganan ketidakseimbangan data.

Kata kunci: XGBoost, Penyakit Jantung, Klasifikasi, Confusion Matrix

ABSTRACT

Heart disease is a serious health challenge that requires accurate prediction and diagnosis. This research explores the application of the eXtreme Gradient Boosting (XGBoost) method in heart disease classification and prediction using patient clinical datasets obtained from Kaggle (<https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>). The focus of the research involved analyzing the impact of XGBoost implementation, the level of accuracy achieved, identification of significant variables, and selection of optimal parameters. This involved comparison with Random Forest and Support Vector Machine (SVM) algorithms. Exploratory Data Analyst described the distribution of characteristics such as age, type of chest pain, blood pressure, and cholesterol in the sample. Feature selection highlights the variables that most influence the output, including exng, cp, oldpeak, and thalachh. The performance evaluation results of the XGBoost model using Confusion Matrix showed good accuracy (85.25%), adequate precision (81.57%), high recall (93.94%), and f1-score of 87.32%. In comparison SVM produced the best value with an accuracy of 86.67% and f1-score of 87.50% outperforming the performance of XGBoost with an accuracy of 85.25% and f1-score of 87.32% and Random Forest with an accuracy of 81.97% and f1-score of 85.34%. The conclusion emphasizes that XGBoost is effective in heart disease classification, and suggestions for future research include implementing the method on other datasets and handling data imbalance.

Keyword: XGBoost, Heart Disease, Classification, Confusion Matrix