

**PERBANDINGAN EFEKTIVITAS HYPERPARAMETER
TUNING PADA ALGORITMA DECISION TREE DAN NAÏVE
BAYES DALAM MEMPREDIKSI PENYAKIT JANTUNG**

SKRIPSI

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



disusun oleh

SRI UTAMI

21.11.3950

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

**PERBANDINGAN EFEKTIVITAS HYPERPARAMETER
TUNING PADA ALGORITMA DECISION TREE DAN NAÏVE
BAYES DALAM MEMPREDIKSI PENYAKIT JANTUNG**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
SRI UTAMI
21.11.3950

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN EFEKTIVITAS HYPERPARAMETER
TUNING PADA ALGORITMA DECISION TREE DAN NAÏVE
BAYES DALAM MEMPREDIKSI PENYAKIT JANTUNG**

yang disusun dan diajukan oleh

Sri Utami

21.11.1998

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 Februari 2025

Dosen Pembimbing,


Komari S. Kusuma M. Eng
NIK. 190302112

HALAMAN PENGESAHAN
SKRIPSI
PERBANDINGAN EFEKTIVITAS HYPERPARAMETER
TUNING PADA ALGORITMA DECISION TREE DAN NAÏVE
BAYES DALAM MEMPREDIKSI PENYAKIT JANTUNG

yang disusun dan diajukan oleh

Sei Utami

21.11.1990

Telah dipertimbangkan di depan Dewan Pengaji
pada tanggal 21 Februari 2025

Susunan Dewan Pengaji

Nama Pengaji

Tanda Tangan

Rumini, S.Kom., M.Kom.
NIK. 190302246

Mutha Sahriyana, S.Kom., M.Kom.
NIK. 190302248

Kurniawati, S.Kom., M.Ts.
NIK. 190302112

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 Februari 2025

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatah, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Sri Utami
NIM : 21.11.3950

Menyatakan bahwa Skripsi dengan judul berikut:

Perbandingan Efektivitas Hyperparameter Tuning pada Algoritma Decision Tree dan Naïve Bayes dalam Memprediksi Penyakit Jantung

Dosen Pembimbing : Kusnawi, S.Koen., M.Eng

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, 21 Februari 2025

Yang Menyatakan,



SRI UTAMI

HALAMAN PERSEMBAHAN

Alhamdulillah hirobbil 'alamin atas berkat dan rahmat serta kemudahan dari Allah Subhanahu wa Ta'ala karya ini penulis persembahkan untuk:

Allah Subhanahu Wa Ta'ala

Puji syukur kehadiran Allah SWT yang telah memberikan rahmat-Nya berupa kesehatan, kemudahan dan kesempatan sehingga peneliti dapat menyelesaikan skripsi ini.

Orang Tua dan Kedua Kakak

Terimakasih kepada bapak, mama dan kedua kakak yang tercinta yang selalu memberikan dukungan, memberikan cinta, mendoakan dan memberikan semangat sehingga penulis dapat menyelesaikan penelitian ini dengan baik. Penulis sangat bangga memiliki keluarga yang selalu mendukung dan penuh cinta.

KATA PENGANTAR

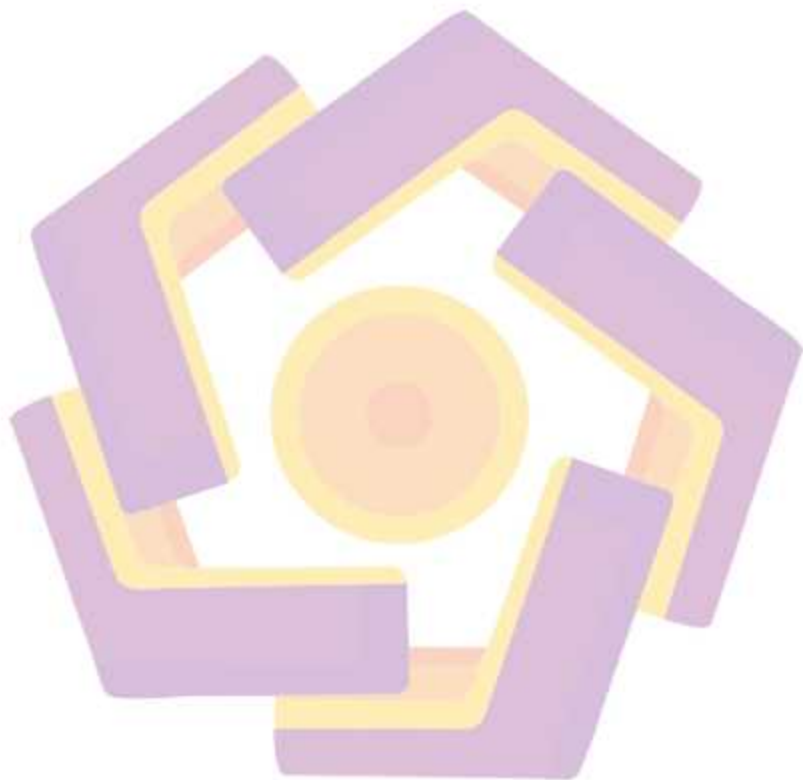
Puji syukur penulis panjatkan ke hadirat Allah SWT yang telah memberikan rahmat-Nya yang berupa kesehatan, karunia, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi ini. Penulis menyadari bahwa dalam penyelesaian penyusunan skripsi ini banyak pihak yang terlibat, sehingga penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua tercinta dan kedua kakak, yang selalu memberikan dukungan, cinta, serta doa. Terimakasih atas kesabaran, semangat dan motivasi yang tiada pernah surut. Semua pencapaian ini tak lepas dari kasih sayang dan dukungan yang luar biasa.
2. Bapak Prof. Dr. Mohammad Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.
3. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D., selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Ibu Windha Mega Pradnya Dhuhiya, S.Kom., M.Kom., selaku Ketua Program Studi S1 Informatika Universitas Amikom Yogyakarta.
5. Bapak Kusnawi, S.Kom., M.Eng. selaku Dosen pembimbing, yang dengan tulus dan sabar telah bersedia dan meluangkan waktu dan tenaganya untuk memberikan bimbingan, arahan, dukungan, saran, dan ilmu yang sangat berarti hingga penulisan skripsi ini dapat terselesaikan.
6. Tim Dosen Penguji, atas kritik, saran dan evaluasi yang membangun untuk menyempurnakan skripsi ini.
7. Teman-teman Informatika 02, yang telah memberikan dukungan, semangat dan kebersamaan, baik dalam suka maupun duka selama menjalani masa perkuliahan.

Akhir kata, semoga Allah SWT senantiasa memberikan rahmat dan karuniaNya serta membalas segala amal baik dan kebaikan dari semua pihak yang telah membantu penulis dalam menyelesaikan skripsi ini.

Yogyakarta, 21 Februari 2025

Penulis



DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xv
DAFTAR LAMBANG DAN SINGKATAN.....	xvi
DAFTAR ISTILAH.....	xvii
INTISARI.....	xviii
<i>ABSTRACT</i>	xix
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.....	3
BAB II TINJAUAN PUSTAKA.....	5

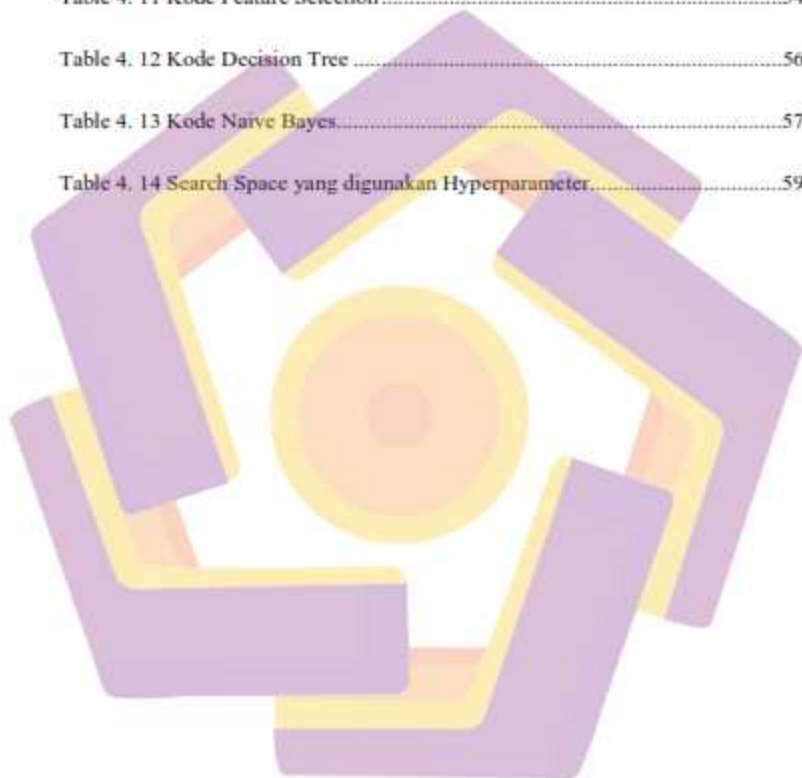
2.1 Studi Literatur.....	5
2.2 Dasar Teori	15
2.2.1 Penyakit Jantung	15
2.2.2 <i>Decision Tree</i>	16
2.2.3 <i>Naive Bayes</i>	17
2.2.4 <i>Hyperparameter Tuning</i>	18
2.2.5 Metode Evaluasi.....	18
BAB III METODE PENELITIAN.....	21
3.1 Objek Penelitian	21
3.2 Alur Penelitian.....	21
3.2.1 Dataset.....	23
3.2.2 Exploratory Data Analysis (EDA).....	24
3.2.3 Preprocessing	26
3.2.4 Classification.....	30
3.2.5 Hyperparameter Tuning dan K-Fold.....	41
3.2.6 Evaluasi.....	42
3.3 Alat dan Bahan	42
BAB IV HASIL DAN PEMBAHASAN	44
4.1 Data Collection.....	44
4.2 Preprocessing.....	48

4.3 Classification	56
4.4 Hyperparameter Tuning dan K-Fold	59
4.5 Hasil Evaluasi Model	62
4.5.1 Classification Report.....	63
4.5.2 Confusion Matrix	65
4.6 Implementasi Website	67
4.7 Analisis Perbandingan Penelitian	69
BAB V PENUTUP	70
5.1 Kesimpulan.....	70
5.2 Saran	70
REFERENSI.....	72
LAMPIRAN	77

DAFTAR TABEL

Table 2. 1 Keaslian Penelitian	8
Table 3. 1 Deskripsi Dataset Heart Disease Predict.....	23
Table 3. 2 Informasi Dataset.....	24
Table 3. 3 Jumlah Dataset Setiap Fitur.....	24
Table 3. 4 Missing Values	26
Tabel 3. 5 Memeriksa Outlier di Semua Variabel	27
Table 3. 6 Imbalanced Data	28
Table 3. 7 Nilai Fitur Corelation with Target	29
Tabel 3. 8 Menampilkan 5 Data Teratas	30
Tabel 3. 9 Entropy dan Gain	34
Tabel 3. 10 Data dalam Kategori	35
Tabel 3. 11 Hasil Classification Naive Bayes.....	41
Table 4. 1 Deskripsi Type Data pada Fitur dalam Dataset	44
Table 4. 2 Kode Scatter Plot	45
Table 4. 3 Kode Mendeteksi Outlier.....	47
Table 4. 4 Kode Missing Values.....	48
Table 4. 5 Kode IQR dan Winsorizing	49
Table 4. 6 Kode Renaming Columns.....	50
Tabel 4. 7 After Renaming Columns	51

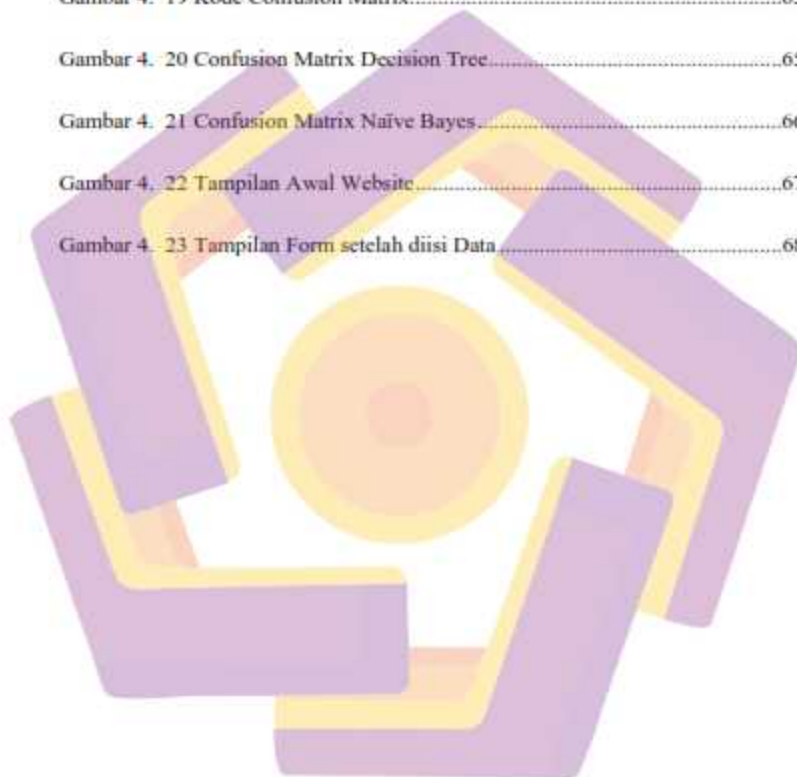
Table 4. 8 Kode Imbalanced Data (SMOTE)	51
Table 4. 9 Kode Split Data.....	53
Table 4. 10 Hasil Split Data.....	54
Table 4. 11 Kode Feature Selection	54
Table 4. 12 Kode Decision Tree	56
Table 4. 13 Kode Naive Bayes.....	57
Table 4. 14 Search Space yang digunakan Hyperparameter.....	59



DAFTAR GAMBAR

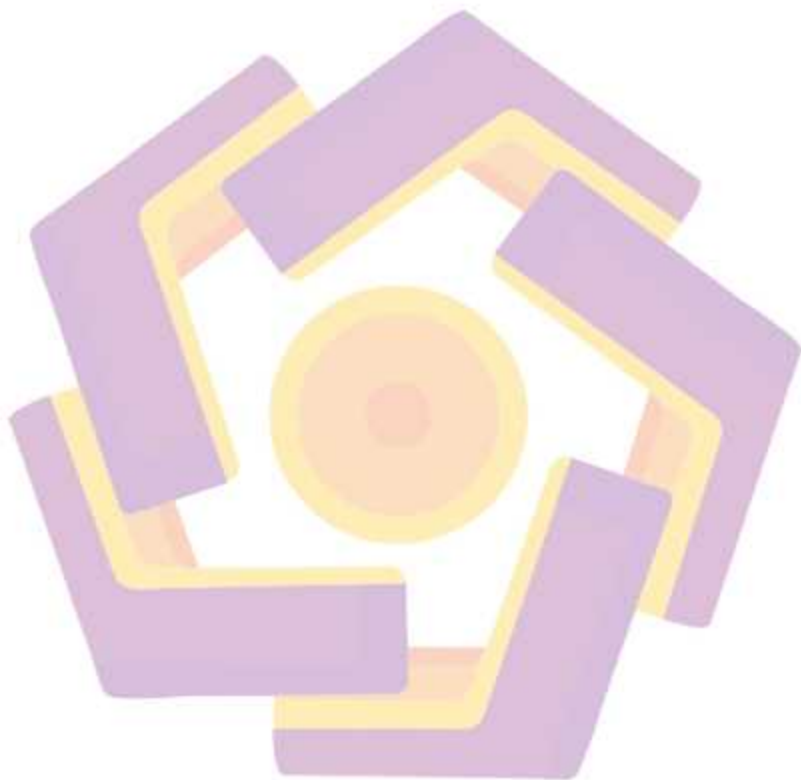
Gambar 2. 1 Struktur Decision Tree	16
Gambar 2.2 Confusion Matrix	19
Gambar 3. 1 Alur Penelitian	22
Gambar 3. 2 Root Node	35
Gambar 4. 1 Fitur sysBP vs diaBP (dengan kategori TenYearCHD).....	45
Gambar 4. 2 Fitur age vs sysBP (dengan kategori TenYearCHD).....	46
Gambar 4. 3 Fitur diaBP vs BMI (dengan kategori TenYearCHD)	46
Gambar 4. 4 Fitur BMI vs sysBP (dengan kategori TenYearCHD).....	47
Gambar 4. 5 Mendeteksi Outlier pada Semua Fitur.....	48
Gambar 4. 6 Nilai Null pada Dataset	49
Gambar 4. 7 Hasil Winsorizing.....	50
Gambar 4. 8 Distribusi Target Sebelum SMOTE.....	52
Gambar 4. 9 Distribusi Target Setelah SMOTE	53
Gambar 4. 10 Feature Selection Berdasarkan Korelasi	55
Gambar 4. 11 Classification Report Decision Tree	56
Gambar 4. 12 Confusion Matrix Decision Tree.....	57
Gambar 4. 13 Classification Report Naive Bayes.....	58
Gambar 4. 14 Confusion Matrix Naive Bayes	59
Gambar 4. 15 Pohon Keputusan Decision Tree	62

Gambar 4. 16 Kode Classification Report	63
Gambar 4. 17 Classification Report Decision Tree	64
Gambar 4. 18 Classification Report Naïve Bayes.....	64
Gambar 4. 19 Kode Confusion Matrix.....	65
Gambar 4. 20 Confusion Matrix Decision Tree.....	65
Gambar 4. 21 Confusion Matrix Naïve Bayes.....	66
Gambar 4. 22 Tampilan Awal Website.....	67
Gambar 4. 23 Tampilan Form setelah diisi Data.....	68



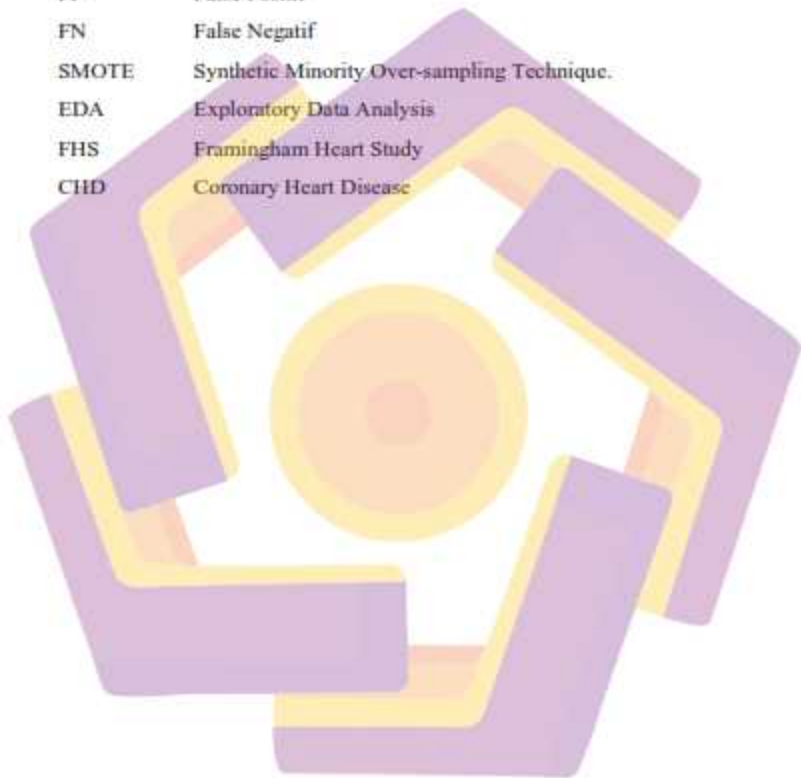
DAFTAR LAMPIRAN

Lampiran 1. Kode Program Python	77
Lampiran 2. Kode Implementasi Website	86



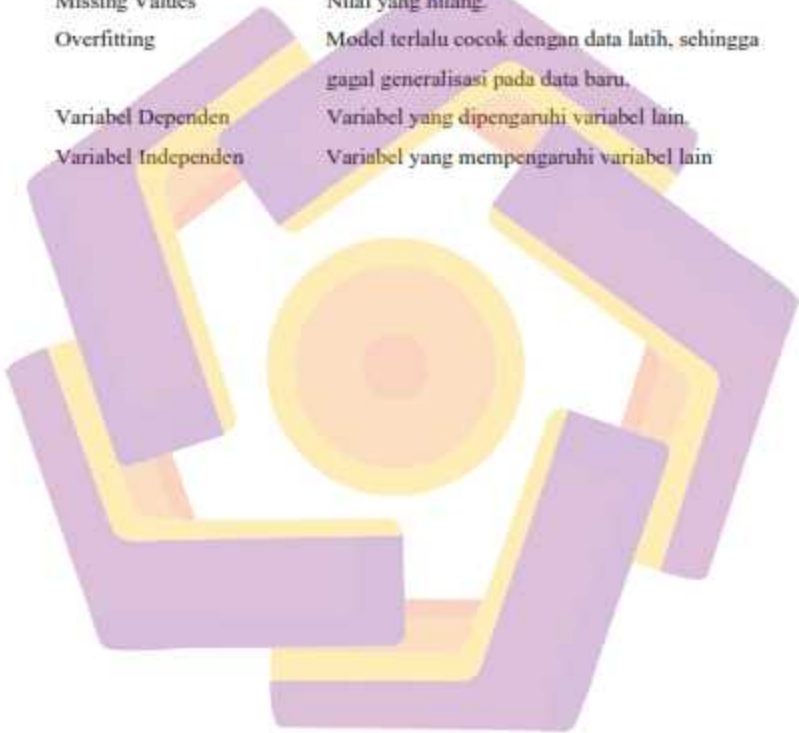
DAFTAR LAMBANG DAN SINGKATAN

IQR	Interquartile Range
TP	True Positif
TN	True Negatif
FP	False Positif
FN	False Negatif
SMOTE	Synthetic Minority Over-sampling Technique.
EDA	Exploratory Data Analysis
FHS	Framingham Heart Study
CHD	Coronary Heart Disease



DAFTAR ISTILAH

Winsorizing	Mengganti nilai dengan batas.
Hyperparameter Tuning	Proses menyesuaikan parameter untuk meningkatkan kinerja model.
Random Search	Metode pencarian hyperparameter secara acak.
Missing Values	Nilai yang hilang.
Overfitting	Model terlalu cocok dengan data latih, sehingga gagal generalisasi pada data baru.
Variabel Dependen	Variabel yang dipengaruhi variabel lain.
Variabel Independen	Variabel yang mempengaruhi variabel lain.



INTISARI

Jantung merupakan penyebab kematian tertinggi di dunia maupun di Indonesia. Pada tahun 2021, WHO mencatat lebih dari 17,9 juta orang meninggal karena penyakit kardiovaskular[1]. Angka kematian standar usia untuk penyakit kardiovaskular (CVD) di Indonesia meningkat dari 356,05 menjadi 412,46 kematian dari tahun 2000 hingga 2019[2]. Deteksi dini menjadi langkah penting untuk meminimalisir risiko yang dapat membahayakan pasien. Penelitian ini bertujuan untuk membandingkan efektivitas *hyperparameter tuning* pada algoritma *Decision Tree* dan *Naïve Bayes* dalam memprediksi penyakit jantung menggunakan dataset *Framingham Heart Study*.

Tahapan penelitian ini melakukan beberapa tahapan *preprocessing* data meliputi: penanganan nilai yang hilang, deteksi serta penanganan *outlier*, dan teknik *Synthetic Minority Oversampling Technique (SMOTE)* untuk menyeimbangkan data yang tidak proporsional. *Hyperparameter tuning* diterapkan dengan menggunakan metode *Random Search* dan *K-Fold Cross-Validation* guna meningkatkan kinerja model. Evaluasi dilakukan menggunakan *Classification Report* dan *Confusion Matrix* berdasarkan metrik akurasi, presisi, dan *recall*.

Hasil penelitian menunjukkan bahwa algoritma *Decision tree* mampu mencapai akurasi terbaik bahkan sebelum di terapkan *tuning*, dengan akurasi 74% menjadi 84%, maka dapat dilihat terdapat peningkatan yang signifikan sebesar 10%. Sedangkan *Naïve Bayes* memiliki peningkatan akurasi 1%, dari 63% menjadi 64%. Penelitian di masa depan disarankan untuk mengeksplor algoritma yang lain dan menggunakan dataset yang lebih banyak.

Kata kunci: Penyakit Jantung, Hyperparameter Tuning, Decision Tree, Naïve Bayes, Klasifikasi.

ABSTRACT

Heart disease is the highest cause of death in the world and in Indonesia. In 2021, the WHO recorded that more than 17.9 million people died from cardiovascular disease[1]. The age-standardized mortality rate for cardiovascular disease (CVD) in Indonesia increased from 356.05 to 412.46 deaths from 2000 to 2019[2]. Early detection is a crucial step to reduce risks that can harm patients. This study compares the effectiveness of hyperparameter tuning in the Decision Tree and Naïve Bayes algorithms in predicting heart disease using the Framingham Heart Study dataset.

The stages of this research include data preprocessing, handling missing values, outlier detection and control, and the Synthetic Minority Oversampling Technique (SMOTE) to balance disproportionate data. Hyperparameter tuning is applied using the Random Search method to improve model performance. The evaluation uses cross-validation and a confusion matrix based on accuracy, precision, and recall metrics.

The results showed that the Decision tree algorithm achieved the best accuracy of 84% after parameter tuning, higher than Naïve Bayes, which reached 64%. Future research is recommended to explore other algorithms and use more datasets.

Keywords: Heart Disease, Hyperparameter Tuning, Decision Tree, Naïve Bayes, Classification.