

**OPTIMASI HYPERPARAMETER GRIDSEARCHCV PADA
MACHINE LEARNING UNTUK KLASIFIKASI PENYAKIT
JANTUNG**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

SEPTA MIFTAKHUL JANNAH

21.11.3916

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**OPTIMASI HYPERPARAMETER GRIDSEARCHCV PADA
MACHINE LEARNING UNTUK KLASIFIKASI PENYAKIT
JANTUNG**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

SEPTA MIFTAKHUL JANNAH

21.11.3916

Kepada

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

HALAMAN PERSETUJUAN

SKRIPSI

**OPTIMASI HYPERPARAMETER GRIDSEARCHCV PADA
MACHINE LEARNING UNTUK KLASIFIKASI PENYAKIT
JANTUNG**

yang disusun dan diajukan oleh

Septa Miftakhul Jannah

21.11.3916

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 23 Desember 2024

Dosen Pembimbing,


Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393

HALAMAN PENGESAHAN
SKRIPSI
OPTIMASI HYPERPARAMETER GRIDSEARCHCV PADA
MACHINE LEARNING UNTUK KLASIFIKASI PENYAKIT
JANTUNG

yang disusun dan diajukan oleh

Septa Miftakhul Jannah

21.11.3916

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Desember 2024

Susunan Dewan Penguji

Nama Penguji

Muhammad Rudyanto Arief, S.T., M.T.
NIK. 190302098

Mulia Sulistiyono, S.Kom., M.Kom.
NIK. 190302248

Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393

Tanda Tangan



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

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Septa Miftakhul Jannah
NIM : 21.11.3916

Menyatakan bahwa Skripsi dengan judul berikut:

Tuliskan Judul Skripsi

Dosen Pembimbing : Majid Rahardi, S.Kom., 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, 23 Desember 2024

Yang Menyatakan,



Septa Miftakhul Jannah

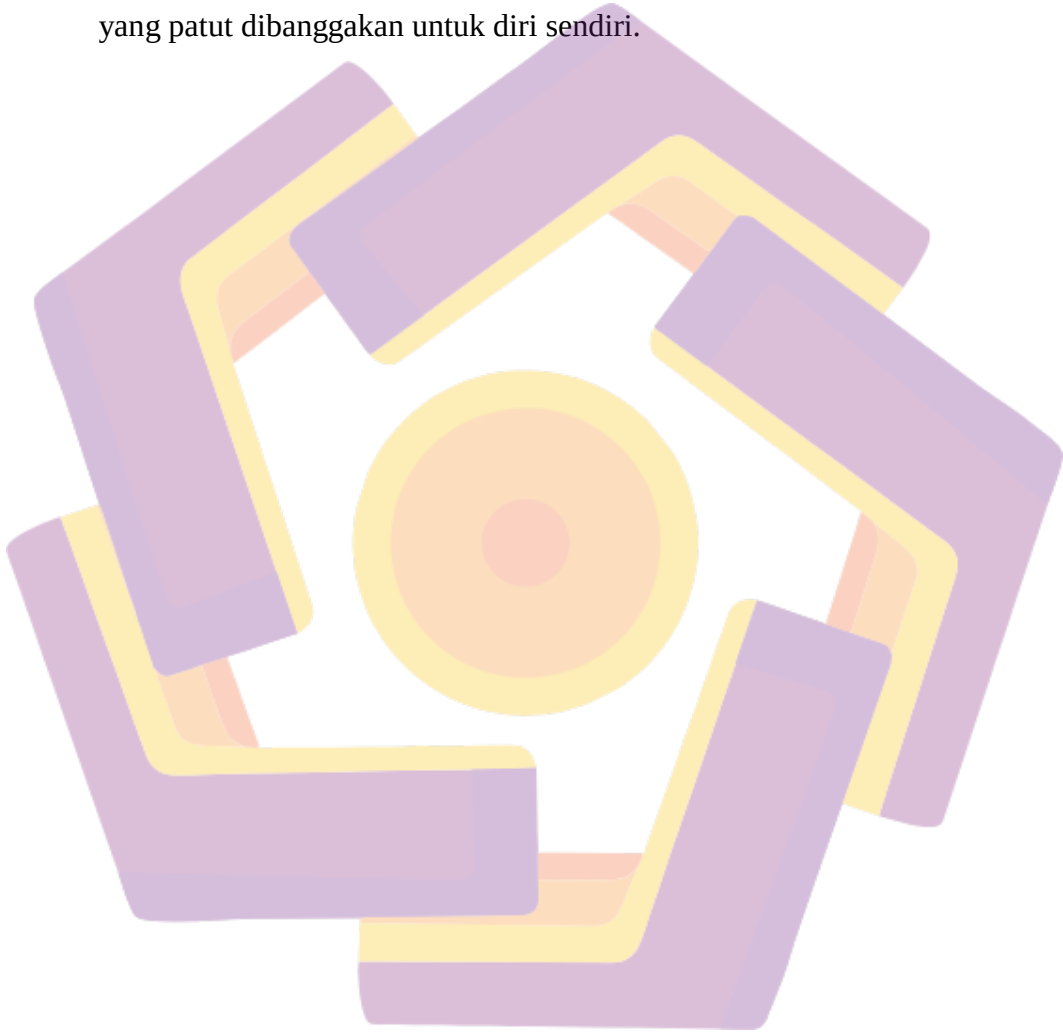
HALAMAN PERSEMBAHAN

Segala puji dan syukur saya persembahkan kepada Allah subhānahu wa ta'āla yang selalu memberikan rahmat dan karunia-Nya sehingga saya dapat menyelesaikan naskah skripsi ini dengan lancar dan barokah. Naskah skripsi ini saya persembahkan kepada:

1. Kedua orang tua tercintaku, bapak Sarimin Fauzie dan ibu Eni Ruswita yang menjadi alasan utama untuk penulis dapat bertahan dalam setiap proses yang penulis jalani selama perkuliahan. Sebagai wujud jawaban dan tanggung jawab atas kepercayaan yang telah diamankan kepadaku, terima kasih atas cinta dan kasih sayang, kesabaran yang tulus membesarkan, merawat dan memberikan do'a, dukungan moral maupun material. Pengorbanan kalian yang tidak ternilai harganya telah mengantarkan penulis menyelesaikan studi S1 di Universitas Amikom Yogyakarta. Kebahagiaan dan rasa bangga kalian menjadi tujuan utama hidupku.
2. Saudara – saudara tercinta penulis Kakak Rofika Ika Putri, Anggi Dwi Pratiwi, Azzahra Munawaroh, Abang Fakhil Fajar Ardiansyah, Thoriq Catur Khusnanto dan kedua adek penulis Nabila Zulva Khafidoh dan M. Raihan Firdaus penulis ucapkan terimakasih karena selalu mendoakan dan memberikan semangat kepada penulis dan juga selalu memberikan bantuan baik material maupun non-material serta selalu ada disaat suka maupun duka.
3. Universitas Amikom Yogyakarta, terima kasih telah menjadi tempat saya menimba ilmu dan melanjutkan studi hingga saya dapat menyelesaikan pendidikan ini.
4. Bapak dosen pembimbing, terima kasih atas bimbingan, arahan, dan motivasi yang sangat berarti dalam menyelesaikan naskah skripsi ini.
5. Sahabat penulis Imroatun Nurul Jannah, Febrian Ainun Nisa, Imsagi Hendi Setiawan dan Rifki Fahrezi Putra I, terimakasih karena telah memberikan semangat, dukungan, dan bantuan kepada penulis selama mengerjakan skripsi ini dan telah mendengarkan keluh kesah penulis.
6. Teman- teman kost bu samsul yang telah kebersamaan hari-hari penulis, memberikan semangat, mengajak healing penulis dan menciptakan banyak

kesan serta do'a yang mendukung untuk penulis selama ini.

7. Semua pihak yang tidak tercantum namanya penulis ucapkan terimakasih yang sebesar-besarnya atas penyelesaian skripsi ini.
8. Terakhir, terimakasih untuk diri sendiri, karena telah mampu berusaha keras dan berjuang sejauh ini. Mampu mengatur waktu, tenaga, pikiran dengan sangat amat baik sehingga dapat mengerjakan skripsi ini dengan menyelesaikan sebaik dan semaksimal mungkin, ini merupakan pencapaian yang patut dibanggakan untuk diri sendiri.



KATA PENGANTAR

Dengan puji syukur saya persembahkan kepada Allah subhānahu wa ta'āla, Tuhan Yang Maha Esa, atas karunia-Nya dan ridho-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Optimasi Hyperparameter GridsearchCV Pada Machine Learning Untuk Klasifikasi Penyakit Jantung” yang mana naskah skripsi ini saya ajukan sebagai salah satu syarat kelulusan pada Program Studi S1 Informatika. Penulis juga mengucapkan terima kasih yang sebesar – besarnya kepada semua pihak yang telah berkontribusi membantu menyelesaikan naskah skripsi ini. Penulis sangat menyadari bahwa tanpa adanya bantuan, arahan dan dukungan dari berbagai pihak maka naskah skripsi ini tidak akan selesai.

Selanjutnya penulis juga mengucapkan terima kasih atas dukungan dan bantuannya sehingga dapat sampai pada tahap ini kepada :

1. Sarimin Fauzie dan Eni Ruswita selaku kedua orang tua yang selalu memberi do'a, dukungan dan semangat kepada penulis.
2. Prof, Dr. M. Suyanto, MM. selaku Rektor Universitas Amikom Yogyakarta.
3. Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer
4. Majid Rahardi, S.Kom., M.Eng selaku Dosen pembimbing yang telah membantu dan memberikan arahan dan bimbingan selama proses pengerjaan skripsi.
5. Segenap Dosen Program Studi S1 Informatika yang telah memberikan wawasan kepada penulis selama proses menimba ilmu di Universitas Amikom Yogyakarta.
6. Serta semua pihak yang selalu memberikan dukungan, doa, motivasi dan semangat sehingga dapat melancarkan dan memudahkan dalam proses pembuatan Skripsi ini.

Yogyakarta, 10 Desember 2024

Penulis

DAFTAR ISI

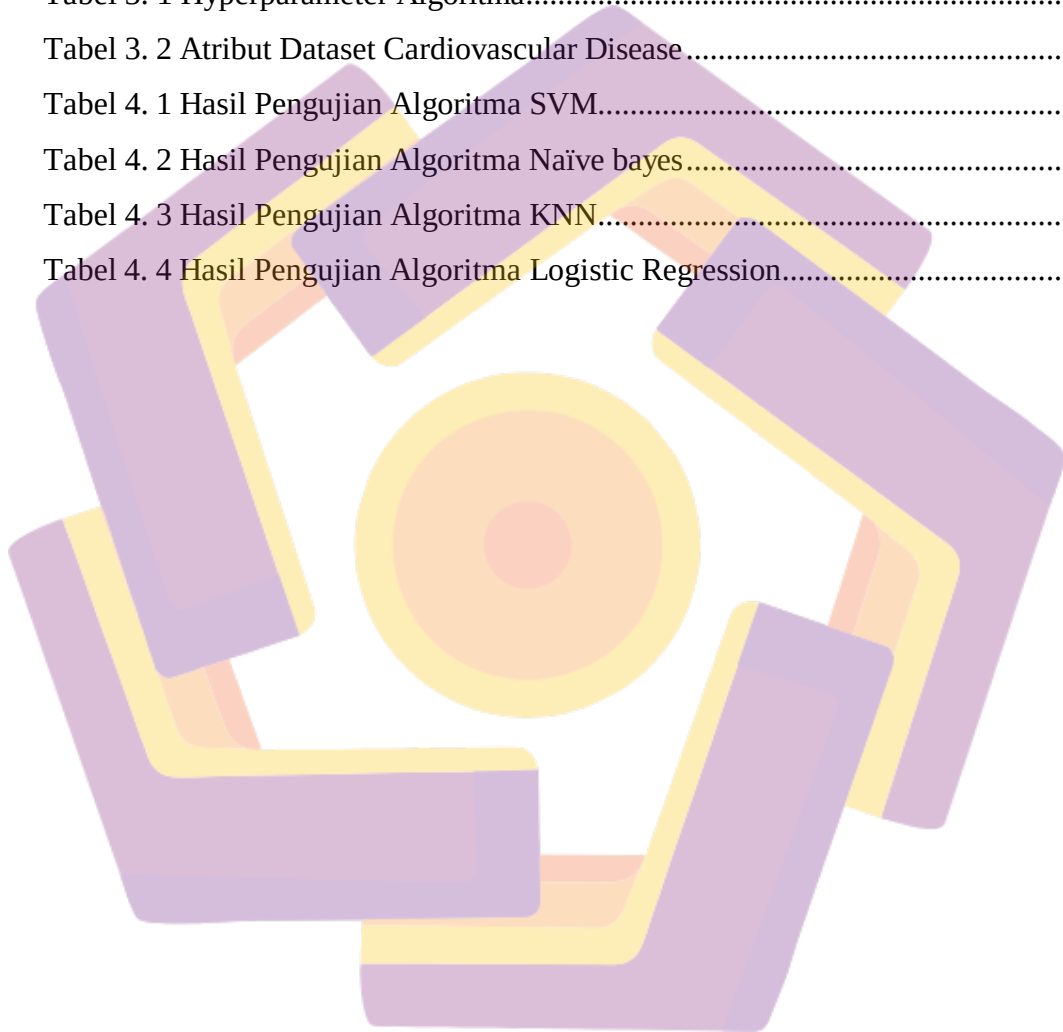
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xvi
ABSTRACT	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori	10
2.2.1 Data Mining	10
2.2.2 Machine Learning	10

2.2.3	Klasifikasi.....	11
2.2.4	Penyakit Jantung.....	12
2.2.5	SMOTE	16
2.2.6	Feature Selection	17
2.2.7	Korelasi Pearson.....	17
2.2.8	Normalisasi data.....	19
2.2.9	Hyperparameter GridsearchCV	19
2.2.10	Algoritma Support Vector Machine	21
2.2.11	Algoritma Naïve Bayes	23
2.2.12	Algoritma Logistic Regression	24
2.2.13	Algoritma K-Nearest Neighbors.....	25
2.2.14	Confusion Matrix	27
BAB III METODE PENELITIAN.....		29
3.1	Objek Penelitian.....	29
3.2	Alur Penelitian	29
3.2.1	Studi literatur.....	30
3.2.2	Memuat dataset.....	30
3.2.3	Preprocessing.....	30
3.2.4	Feature Selection	31
3.2.5	Normalisasi data.....	31
3.2.6	Split data.....	31
3.2.7	Modelling	31
3.2.8	Hyperparameter GridsearchCV	32
3.2.9	Evaluasi model	32
3.3	Alat dan Bahan	33
3.3.1	Data Penelitian.....	33

3.3.2	Alat Penelitian	34
BAB IV HASIL DAN PEMBAHASAN		36
4.1	Memuat dataset.....	36
4.2	Preprocessing data	37
4.3	Korelasi Pearson	40
4.4	Normalisasi data	41
4.5	Split data	42
4.6	Modelling.....	42
4.7	Hyperparameter GridsearchCV	44
4.7.1	Support Vector Machine	44
4.7.2	Naïve Bayes.....	45
4.7.3	K-Nearest Neighbors (KNN).....	46
4.7.4	Logistic Regression	47
4.8	Evaluasi.....	48
4.8.1	Evaluasi Algoritma SVM	48
4.8.2	Evaluasi Algoritma Naïve Bayes.....	51
4.8.3	Evaluasi Algoritma KNN	53
4.8.4	Evaluasi Algoritma Logistic Regression	55
4.9	Hasil.....	58
BAB V PENUTUP		60
5.1	Kesimpulan	60
5.2	Saran	60
REFERENSI		62

DAFTAR TABEL

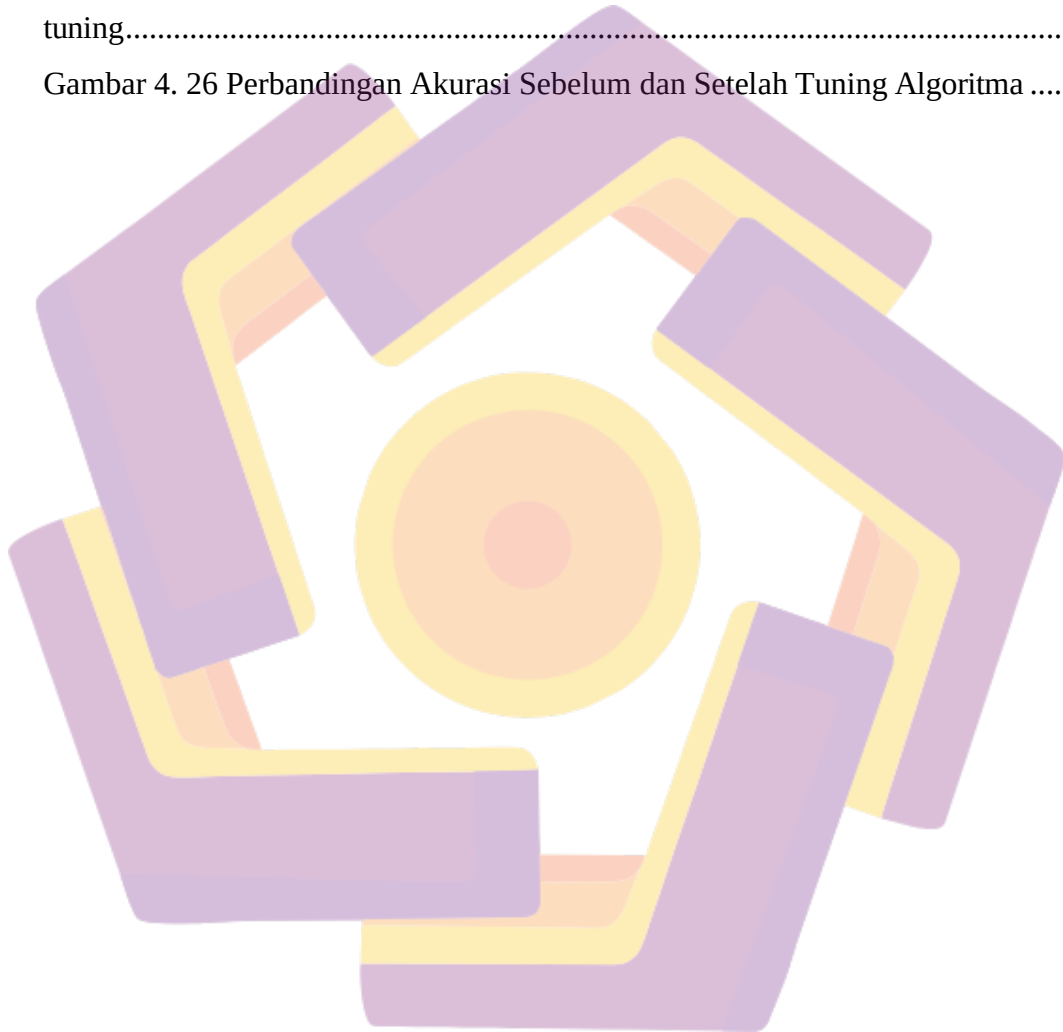
Tabel 2. 1 Keaslian Penelitian	7
Tabel 2. 2 Interpretasi Terhadap Koefisien Korelasi.....	18
Tabel 2. 3 Persamaan Rumus fungsi kernel	22
Tabel 2. 4 Persamaan Rumus Penghitung Performa Klasifikasi	27
Tabel 3. 1 Hyperparameter Algoritma.....	32
Tabel 3. 2 Atribut Dataset Cardiovascular Disease.....	33
Tabel 4. 1 Hasil Pengujian Algoritma SVM.....	50
Tabel 4. 2 Hasil Pengujian Algoritma Naïve bayes.....	52
Tabel 4. 3 Hasil Pengujian Algoritma KNN.....	55
Tabel 4. 4 Hasil Pengujian Algoritma Logistic Regression.....	57



DAFTAR GAMBAR

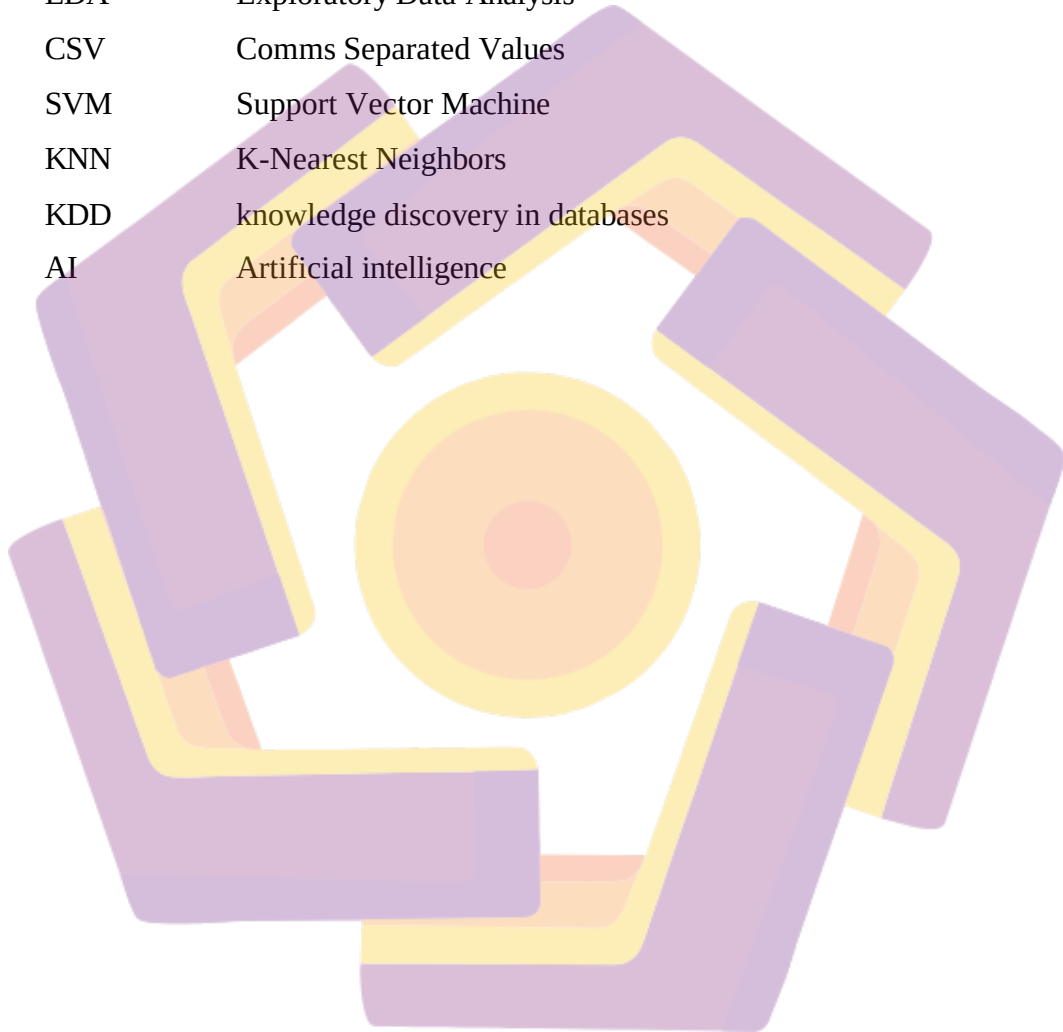
Gambar 2. 1 Ilustrasi <i>hyperplane</i> dan <i>margin</i> maksimal	21
Gambar 2. 2 Transformasi Data	23
Gambar 2. 3 Model naïve bayes	23
Gambar 2. 4 Model Logistic Regression	25
Gambar 2. 5 Model KNN	26
Gambar 3. 1 Alur penelitian	29
Gambar 4. 1 Hasil Pengecekan Tipe data.....	36
Gambar 4. 2 Perbandingan Ada Penyakit Jantung Dan Tidak Ada Penyakit Jantung.....	37
Gambar 4. 3 Hasil Pengecekan Nilai Yang Hilang	38
Gambar 4. 4 Hasil pengecekan duplikat data	38
Gambar 4. 5 Imbalance Data.....	39
Gambar 4. 6 Balance Data.....	40
Gambar 4. 7 Visualisasi Heatmap Korelasi Pearson	41
Gambar 4. 8 Hasil Seleksi 8 Fitur	41
Gambar 4. 9 Hasil Normalisasi Data.....	42
Gambar 4. 10 Classification Report SVM Sebelum Hyperparameter tuning.....	43
Gambar 4. 11 Classification Report Naïve Bayes Sebelum Hyperparameter tuning	43
Gambar 4. 12 Classification Report KNN Sebelum Hyperparameter tuning.....	43
Gambar 4. 13 Classification Report Logistic Regression Sebelum Hyperparameter tuning.....	44
Gambar 4. 14 Classification Report SVM Setelah Hyperparameter tuning.....	45
Gambar 4. 15 Classification Report Naïve Bayes Setelah Hyperparameter tuning	46
Gambar 4. 16 Classification Report KNN Setelah Hyperparameter tuning.....	47
Gambar 4. 17 Classification Report Logistic Regression Setelah Hyperparameter tuning.....	48
Gambar 4. 18 Confusion Matrix SVM Sebelum Hyperparameter tuning.....	50
Gambar 4. 19 Confusion Matrix SVM Setelah Hyperparameter tuning	50

Gambar 4. 20 Confusion Matrix Naïve Bayes Sebelum Hyperparameter tuning..	52
Gambar 4. 21 Confusion Matrix Naïve Bayes Setelah Hyperparameter tuning.....	52
Gambar 4. 22 Confusion Matrix KNN Sebelum Hyperparameter tuning.....	54
Gambar 4. 23 Confusion Matrix KNN Setelah Hyperparameter tuning	55
Gambar 4. 24 Confusion Matrix Logistic Regression Sebelum Hyperparameter tuning.....	57
Gambar 4. 25 Confusion Matrix Logistic Regression Setelah Hyperparameter tuning.....	57
Gambar 4. 26 Perbandingan Akurasi Sebelum dan Setelah Tuning Algoritma	59

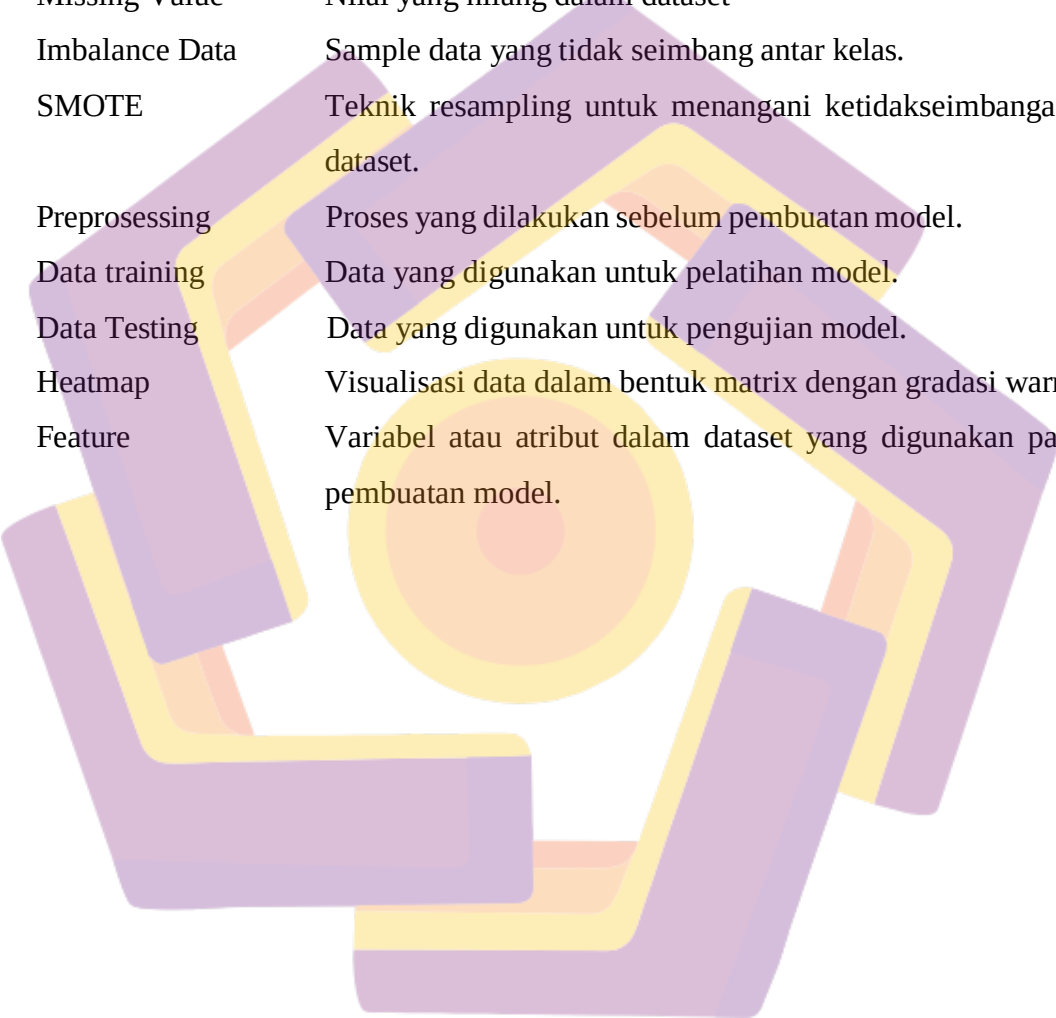


DAFTAR LAMBANG DAN SINGKATAN

WHO	World Health Organization
Kemenkes	Kementrian Kesehatan
ML	Machine Learning
SMOTE	Synthetic Minority Oversampling Technique
EDA	Exploratory Data Analysis
CSV	Comms Separated Values
SVM	Support Vector Machine
KNN	K-Nearest Neighbors
KDD	knowledge discovery in databases
AI	Artificial intelligence



DAFTAR ISTILAH



Variabel Dependen	Variabel yang dipengaruhi variabel lain.
Variabel Independen	Variabel yang mempengaruhi variabel lain.
Dataset	Kumpulan data
EDA	Proses memahami pola dan hubungan pada data.
Cloud	Penyimpanan yang disimpan secara daring.
Missing Value	Nilai yang hilang dalam dataset
Imbalance Data	Sample data yang tidak seimbang antar kelas.
SMOTE	Teknik resampling untuk menangani ketidakseimbangan dataset.
Preprocessing	Proses yang dilakukan sebelum pembuatan model.
Data training	Data yang digunakan untuk pelatihan model.
Data Testing	Data yang digunakan untuk pengujian model.
Heatmap	Visualisasi data dalam bentuk matrix dengan gradasi warna.
Feature	Variabel atau atribut dalam dataset yang digunakan pada pembuatan model.

INTISARI

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia, termasuk di Indonesia. Upaya deteksi dini sangat penting untuk mengurangi risiko kematian, tetapi proses manual sering memakan waktu, biaya, dan rentan kesalahan. Teknologi machine learning menawarkan solusi yang lebih cepat dan efisien dengan menganalisis data kesehatan yang kompleks. Penelitian ini fokus pada optimasi hyperparameter menggunakan GridSearchCV untuk meningkatkan akurasi model klasifikasi penyakit jantung. Beberapa algoritma seperti Support Vector Machine (SVM), Naive Bayes, K-Nearest Neighbors (KNN), dan Logistic Regression digunakan untuk membandingkan hasil sebelum dan setelah optimasi.

Penelitian dimulai dengan preprocessing data, seperti penanganan data tidak seimbang menggunakan SMOTE, seleksi fitur dengan Korelasi Pearson, normalisasi data menggunakan StandardScaler, serta pembagian dataset menjadi data latih dan uji dengan perbandingan 80:20. Setelah itu, dilakukan optimasi hyperparameter dengan GridSearchCV untuk menemukan pengaturan parameter terbaik pada setiap algoritma. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score melalui konfusi matriks.

Hasil penelitian menunjukkan peningkatan akurasi yang signifikan pada semua algoritma setelah optimasi. Akurasi Support Vector Machine (SVM) meningkat dari 52% menjadi 95%, Naive Bayes dari 72% menjadi 78%, Logistic Regression dari 64% menjadi 97%, dan K-Nearest Neighbors (KNN) dari 53% menjadi 96%. Optimasi ini membantu menghasilkan model prediksi yang lebih akurat untuk mendeteksi dini penyakit jantung. Penelitian ini dapat bermanfaat bagi tenaga medis, peneliti, dan pengembang teknologi kesehatan untuk mempercepat dan mempermudah diagnosis penyakit jantung.

Kata kunci: Machine Learning, Penyakit Jantung, GridsearchCV, SMOTE, Klasifikasi.

ABSTRACT

Heart disease is one of the leading causes of death globally, including in Indonesia. Early detection efforts are crucial to reducing mortality risks, but manual processes are often time-consuming, costly, and prone to errors. Machine learning technology offers a faster and more efficient solution by analyzing complex health data. This study focuses on hyperparameter optimization using GridSearchCV to enhance the accuracy of heart disease classification models. Several algorithms, such as Support Vector Machine (SVM), Naive Bayes, K-Nearest Neighbors (KNN), and Logistic Regression, are employed to compare results before and after optimization.

The research begins with data preprocessing, including handling imbalanced data using SMOTE, feature selection with Pearson Correlation, data normalization using StandardScaler, and splitting the dataset into training and testing sets with an 80:20 ratio. Subsequently, hyperparameter optimization is performed with GridSearchCV to identify the best parameter settings for each algorithm. Evaluation is conducted using metrics such as accuracy, precision, recall, and F1-score through a confusion matrix.

The study results indicate a significant improvement in accuracy across all algorithms after optimization. The accuracy of the Support Vector Machine (SVM) increased from 52% to 95%, Naive Bayes from 72% to 78%, Logistic Regression from 64% to 97%, and K-Nearest Neighbors (KNN) from 53% to 96%. This optimization facilitates the development of more accurate predictive models for early heart disease detection. This research can benefit healthcare professionals, researchers, and health technology developers by accelerating and simplifying heart disease diagnosis.

Keywords: Machine Learning, Heart Disease, GridSearchCV, SMOTE, Classification.