

**KOMPARASI AKURASI ALGORITMA SUPPORT VECTOR
MACHINE (SVM), LOGISTIC REGRESSION, DAN NAIVE
BAYES PADA PREDIKSI PENYAKIT JANTUNG**

SKRIPSI

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



disusun oleh

MUHAMMAD DIMAS MUGEN DHUHURI

20.11.3634

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**KOMPARASI AKURASI ALGORITMA SUPPORT VECTOR
MACHINE (SVM), LOGISTIC REGRESSION, DAN NAIVE
BAYES PADA PREDIKSI PENYAKIT JANTUNG**

SKRIPSI

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



disusun oleh

MUHAMMAD DIMAS MUGEN DHUHURI

20.11.3634

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**KOMPARASI AKURASI ALGORITMA SUPPORT VECTOR MACHINE
(SVM), LOGISTIC REGRESSION, DAN NAIVE BAYES PADA PREDIKSI
PENYAKIT JANTUNG**

yang disusun dan diajukan oleh

Muhammad Dimas Mugen Dhuhuri

20.11.3634

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 24 september 2024

Dosen Pembimbing,



Anggit Dwi Hartanto, M.Kom.
NIK. 190302163

HALAMAN PENGESAHAN

SKRIPSI

**KOMPARASI AKURASI ALGORITMA SUPPORT VECTOR MACHINE
(SVM), LOGISTIC REGRESSION, DAN NAIVE BAYES PADA PREDIKSI
PENYAKIT JANTUNG**

yang disusun dan diajukan oleh

Muhammad Dimas Mugen Duhuri

20.11.3634

Telah dipertahankan di depan Dewan Penguji
pada tanggal 29 Juli 2024

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Bayu Setiaji, M.Kom
NIK. 190302216

Anna Baita, M. Kom
NIK. 190302290

Anggit Dwi Hartanto, M.Kom
NIK. 190302163



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 29 Juli 2024

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Muhammad Dimas Mugen Dhuhuri**
NIM : **20.11.3634**

Menyatakan bahwa Skripsi dengan judul berikut:

KOMPARASI AKURASI ALGORITMA SUPPORT VECTOR MACHINE (SVM), LOGISTIC REGRESSION, DAN NAIVE BAYES PADA PREDIKSI PENYAKIT JANTUNG

Dosen Pembimbing : **Anggit Dwi Hartanto, 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, 29 Juli 2024

Yang Menyatakan,



Muhammad Dimas Mugen Dhuhuri

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur dan kerendahan hati, saya persembahkan skripsi ini kepada Allah SWT, Tuhan yang Maha Esa, yang senantiasa memberikan saya kesehatan, kekuatan, dan petunjuk sehingga saya dapat menyelesaikan skripsi ini dengan baik. Segala puji dan syukur hanya kepada-Nya, yang telah memberi saya kesempatan untuk belajar dan berkembang dalam proses ini.

Kepada kedua orang tua tercinta, yang merupakan sumber inspirasi dan motivasi terbesar dalam hidup saya. Terima kasih atas cinta, doa, dan pengorbanan yang tidak ternilai harganya. Tanpa dukungan dan pengorbanan kalian, saya tidak akan berada di titik ini. Kalian adalah alasan utama saya untuk selalu berusaha memberikan yang terbaik.

Kepada dosen pembimbing saya, yang telah dengan sabar membimbing dan memberikan arahan selama proses penelitian dan penulisan skripsi ini. Terima kasih atas waktu, pengetahuan, dan nasihat yang sangat berharga. Bimbingan dan dedikasi Anda menjadi pendorong besar bagi saya dalam menyelesaikan tugas akhir ini.

Untuk para pengajar dan dosen di universitas ini, terima kasih atas ilmu dan wawasan yang telah kalian berikan selama masa studi. Setiap pelajaran dan pengalaman yang saya dapatkan sangat berkontribusi dalam membentuk pemahaman dan keterampilan saya.

Kepada teman-teman seperjuangan, terima kasih atas kebersamaan dan dukungan moral yang kalian berikan. Bersama kalian, setiap tantangan menjadi lebih mudah dihadapi, dan setiap pencapaian menjadi lebih berarti untuk dirayakan.

Terakhir, untuk almamater saya, yang telah menjadi rumah kedua selama beberapa tahun terakhir. Semoga skripsi ini dapat menjadi kontribusi kecil yang bermanfaat bagi perkembangan ilmu pengetahuan di bidang yang saya tekuni dan menjadi inspirasi bagi mahasiswa lain yang akan melanjutkan perjalanan ini.

KATA PENGANTAR

Puji syukur kehadirat ALLAH SWT karena berkat rahmat serta hidayah-Nya penulis diberikan kesehatan dan kekuatan serta iman yang kuat sehingga dapat menyelesaikan skripsi yang berjudul “KOMPARASI AKURASI ALGORITMA SUPPORT VECTOR MACHINE (SVM), LOGISTIC REGRESSION, DAN NAIVE BAYES PADA PREDIKSI PENYAKIT JANTUNG”. Oleh karena itu penulis mengucapkan syukur karena dapat menyelesaikan skripsi ini. Skripsi ini diajukan kepada Program Studi S1 Informatika, Fakultas Ilmu Komputer, serta Universitas AMIKOM Yogyakarta.

Penulis dalam menyelesaikan skripsi mendapatkan beberapa hambatan dalam berbagai hal, namun banyak pihak yang selalu membantu sehingga skripsi ini dapat diselesaikan dengan baik. Dengan begitu penulis hendak mengungkapkan terima kasih terhadap:

1. Bapak prof. Dr. M Suyanto, M.M selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Windha Mega Pradya D, M.Kom., selaku Ketua Program Studi S1 Informatika Universitas AMIKOM Yogyakarta.
4. Bapak Anggit Dwi Hartanto, M.Kom selaku dosen pembimbing yang selalu memberikan saran dan arahan dalam proses penulisan skripsi ini.
5. Bapak Rochmad dan Ibu Karsinah selaku orang tua yang selalu memberikan doa dan motivasi kepada penulis.
6. Sahabat sahabat saya: Ulin, Ageng, Sandi, Faris, Tyas, Joko, Nathan, Teddy, Dama, Ardi, Muiz, Icik, Iqbal, Ahmed, Akmal, Midun, Adi, Hijra, Ririn, Sindhi, Zalza, Alan, tobit, Rafli, Khuswatun, Topik, Faiq, Ali Imron, Reni.

Yogyakarta, <tanggal bulan tahun>

Penulis

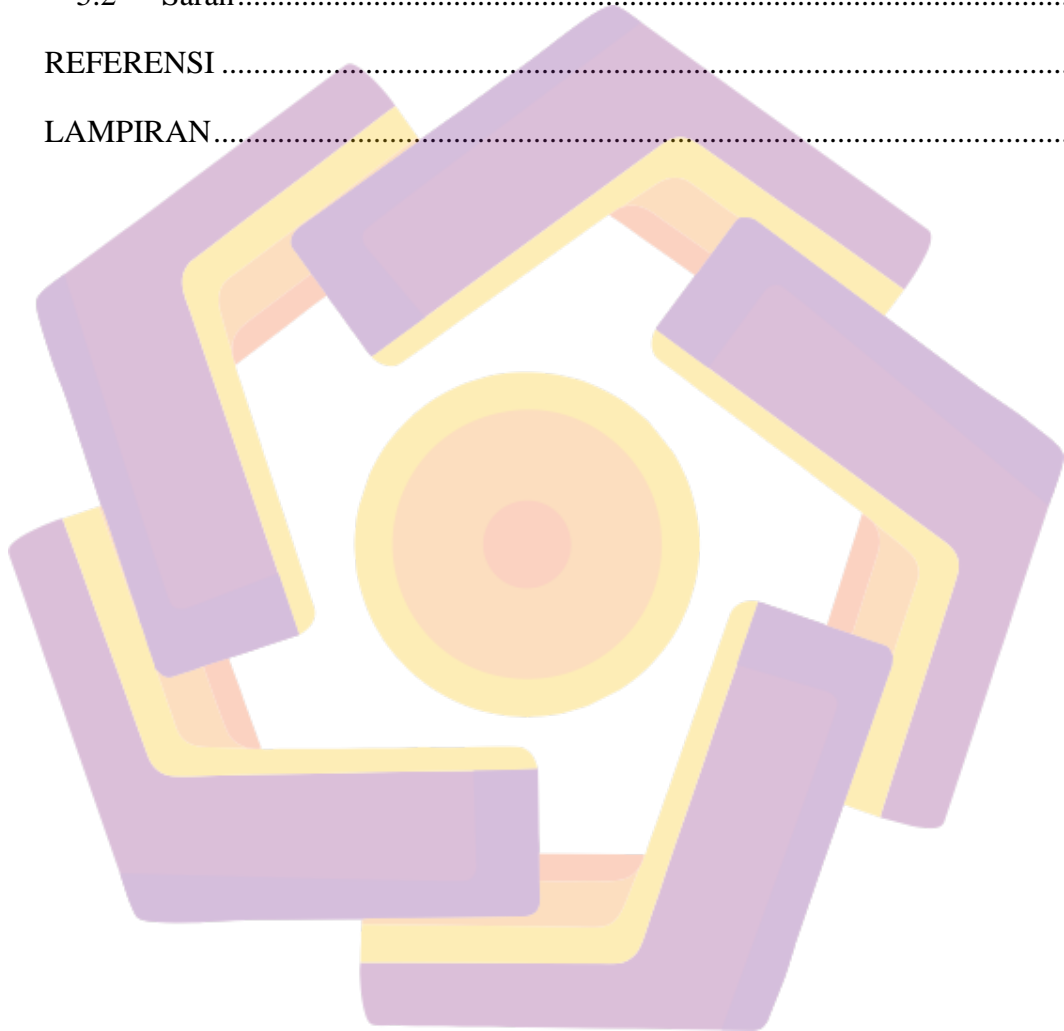
DAFTAR ISI

Contents

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	2
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	3
BAB II TINJAUAN PUSTAKA	4

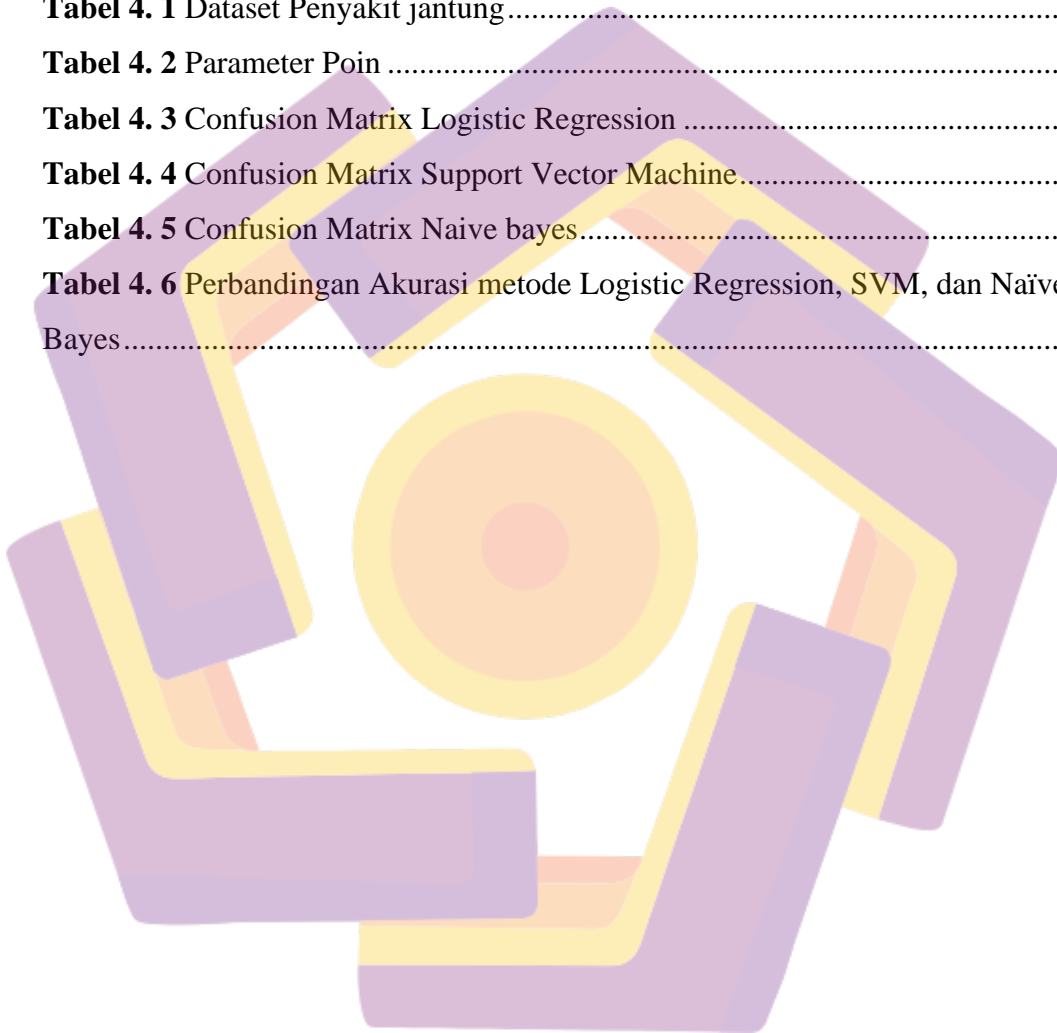
2.1	Studi Literatur	4
2.2	Dasar Teori	8
2.2.1	Penyakit Jantung	8
2.2.2	Komparasi	8
2.2.3	Machine Learning	8
2.2.4	Support Vector Machine	9
2.2.5	Logistic Regression	10
2.2.6	Naïve Bayes	11
2.2.7	Confusion Matrix	11
BAB III METODE PENELITIAN		14
3.1	Alur Penelitian	14
3.2	Pengumpulan Data	14
3.3	Pengolahan data	17
3.4	Tahap Analisis Data	18
3.5	Alat dan Bahan	21
BAB IV HASIL DAN PEMBAHASAN		23
4.1	Pengambilan Data	23
4.2	Load Data	23
4.3	Exploratory Data Analysis (EDA)	23
4.4	Feature Selection	27
4.5	Evaluasi Feature Selection	29
4.6	Training Dataset 80%	29
4.7	Testing Dataset 20%	29
4.8	Logistic Regression	30
4.9	Support Vector Machine	30

4.10	Naïve Bayes	30
4.11	Evaluasi.....	30
BAB V PENUTUP		37
5.1	Kesimpulan.....	37
5.2	Saran.....	37
REFERENSI		38
LAMPIRAN.....		42



DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian.....	6
Tabel 2. 2 Ilustrasi confusion Matrix	12
Tabel 3. 1 Penjelasan Variabel Penelitian	15
Tabel 4. 1 Dataset Penyakit jantung.....	24
Tabel 4. 2 Parameter Poin	28
Tabel 4. 3 Confusion Matrix Logistic Regression	30
Tabel 4. 4 Confusion Matrix Support Vector Machine.....	31
Tabel 4. 5 Confusion Matrix Naive bayes.....	32
Tabel 4. 6 Perbandingan Akurasi metode Logistic Regression, SVM, dan Naïve Bayes.....	34

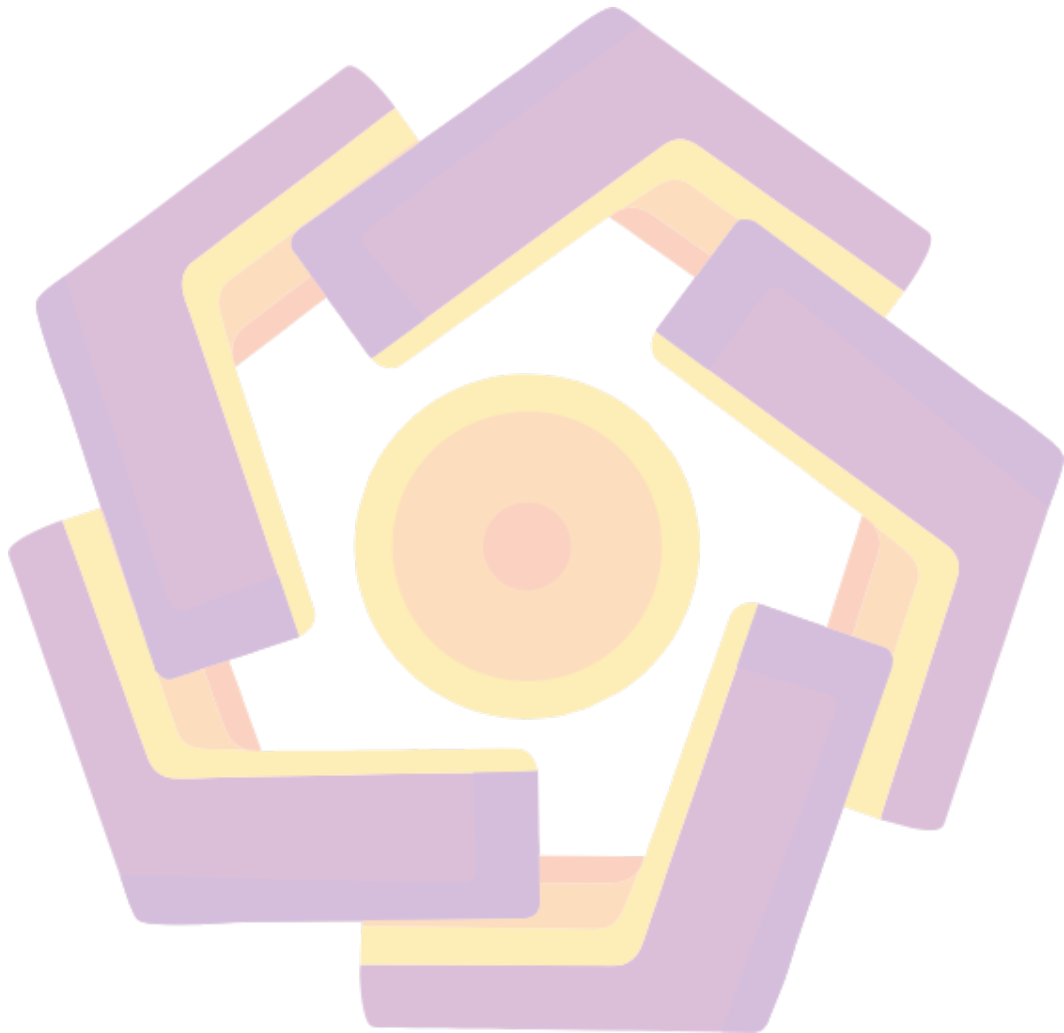


DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	14
Gambar 3. 2 Variabel-Variabel Penelitian	16
Gambar 4. 1 Tampilan Dataset.....	23
Gambar 4. 2 Barplot Distribusi Usia yang Mengalami Penyakit Jantung.....	24
Gambar 4. 3 Barplot Distribusi Usia yang Tidak Mengalami Penyakit Jantung .25	
Gambar 4. 4 Barplot Distribusi Tipe Nyeri pada Dada.....	25
Gambar 4. 5 Barplot Distribusi Tekanan Darah saat Darah Istirahat.....	26
Gambar 4. 6 Barplot Kolesterol.....	27
Gambar 4. 7 Heatmap Feature Selection Dataset.....	28
Gambar 4. 8 Hasil Accuracy, Precision, Recall, dan F1-score Logistic Regression	31
Gambar 4. 9 Hasil Accuracy, Precision, Recall, dan F1-score SVM.....	32
Gambar 4. 10 Hasil Accuracy, Precision, Recall, dan F1-score Naïve Bayes	33
Gambar 4. 11 Perbandingan Akurasi antara Logistic Regression, SVM, dan Naive Bayes	34
Gambar 4. 12 Komparasi SVM sebelum dan sesudah Improvisasi	35

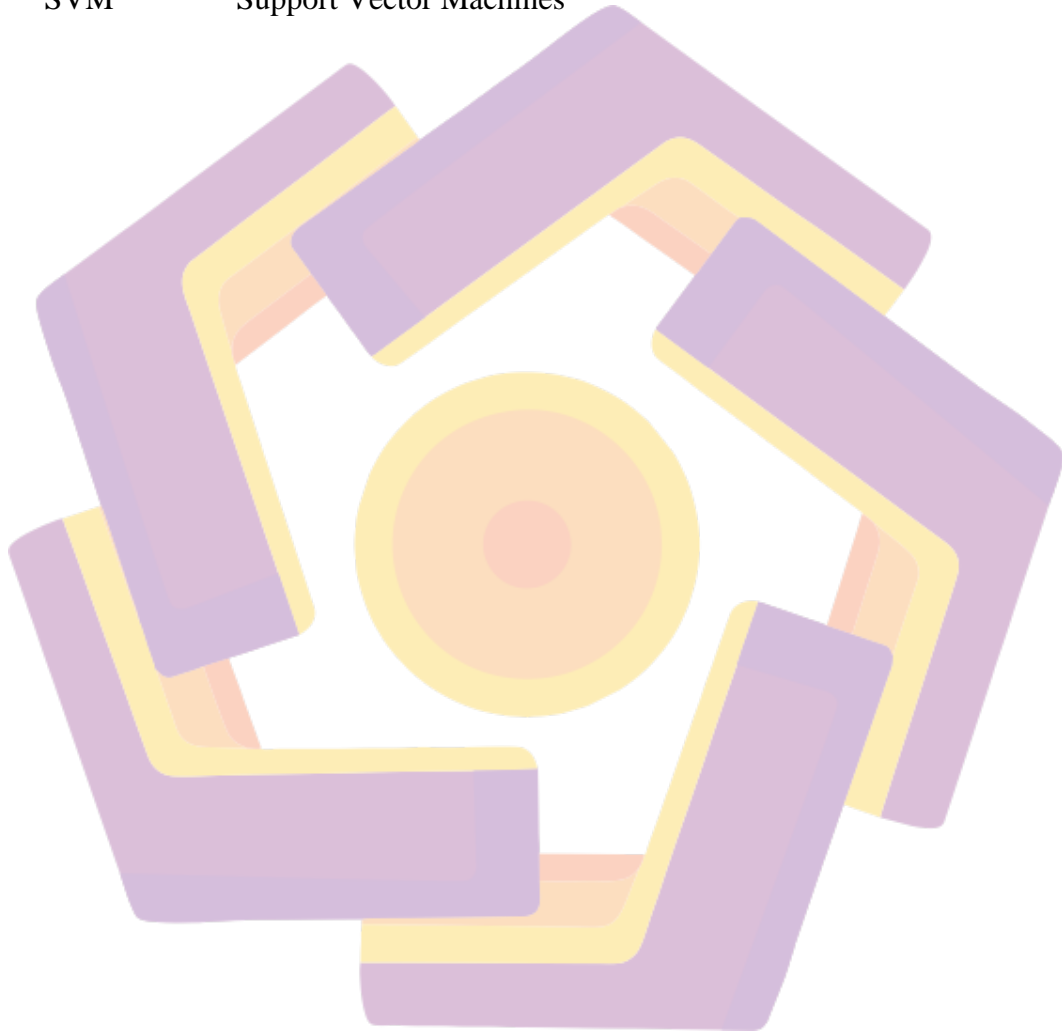
DAFTAR LAMPIRAN

Lampiran 1. Profil obyek Penelitian	10
Lampiran 2. Dokumentasi Penelitian	11



DAFTAR LAMBANG DAN SINGKATAN

Ω	Tahanan Listrik
μ	Konstanta gesekan
ANFIS	Adaptive Network Fuzzy Inference System
SVM	Support Vector Machines



DAFTAR ISTILAH

Dataset	Kumpulan data yang yang biasa digunakan untuk penelitian
Algoritma	instruksi untuk menyelesaikan suatu masalah.
Logistic Regression	Metode statistik untuk memprediksi hasil satu atau lebih variabel independent.
SVM	Metode pembelajaran mesin yang digunakan untuk klasifikasi dan regresi .
Naïve Bayes	Kelompok algoritma klasifikasi sederhana berdasarkan teorema Bayes.
Machine learning	Cabang kecerdasan buatan berfokus pengembangan algoritma dan model.
Heatmap	Representasi grafis dari data ditunjukkan dengan warna.
Confusion Matrix	Tabel evaluasi kinerja model klasifikasi

INTISARI

Penelitian ini bertujuan untuk komparasi/membandingkan performa dari tiga dalam prediksi penyakit jantung, yaitu Support Vector Machine (SVM), Logistic Regression, dan Naïve Bayes. Fokus pada evaluasi yang digunakan untuk mengukur performa algoritma, seperti akurasi, presisi, recall, dan F1-score dan faktor rendahnya nilai akurasi pada algoritma. Dalam klasifikasi dan prediksi penyakit jantung menggunakan dataset yang didapatkan dari Kaggle <https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset> yang pernah digunakan untuk penelitian sebelumnya. Exploratory Data Analyst menunjukkan distribusi karakteristik seperti usia, jenis nyeri dada, tekanan darah, dan kolesterol pada sampel. Pada Feature selection mengamati variabel yang paling mempengaruhi output, termasuk exng, cp, oldpeak, dan thalachh. Hasil Evaluasi ketiga model Logistic Regression, Support Vector Machine (SVM), dan Naïve Bayes menggunakan Confusion Matrix Menunjukkan Logistic Regression mendapatkan nilai akurasi paling tinggi dengan persentase (88,52%) diikuti dengan Naïve Bayes dengan persentase (83,52%) dan yang terakhir Support Vector Machine dengan persentase (57,38%). Terdapat hasil akurasi yang rendah pada Support vector Machine (SVM) dilakukan improvisasi dengan membuat percobaan baru di model SVM untuk mendapatkan parameter terbaik, disini dibantu dengan gridsearchCV, dengan pilihan parameter C, kernel, degree, dan gamma. Dimana C merupakan kelipatan perkalian 10, gamma kelipatan 2, dan kernel kelipatan 3, Lalu menghasilkan parameter dengan akurasi 81,32% dan F1-score sebesar 84,21%

Kata kunci: Penyakit Jantung, Support Vector Machine, Logistic Regression, Naïve Bayes, Prediksi.

ABSTRACT

This study aims to compare the performance of three heart disease prediction algorithms, namely Support Vector Machine (SVM), Logistic Regression, and Naïve Bayes. The focus is on the evaluation used to measure the performance of the algorithms, such as accuracy, precision, recall, and F1-score and the factor of low accuracy value in the algorithm. In the classification and prediction of heart disease, a dataset obtained from Kaggle <https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset> was used for previous research. Exploratory Data Analyst shows the distribution of characteristics such as age, type of chest pain, blood pressure, and cholesterol in the sample. Feature selection looked at the variables that most influenced the output, including exng, cp, oldpeak, and thalachh. The results of evaluating the three models of Logistic Regression, Support Vector Machine (SVM), and Naïve Bayes using Confusion Matrix show that Logistic Regression has the highest accuracy value with a percentage of (88.52%) followed by Naïve Bayes with a percentage of (83.52%) and finally Support Vector Machine with a percentage of (57.38%). There are low accuracy results on Support vector Machine (SVM) improvised by making new experiments in the SVM model to get the best parameters, here assisted by gridsearchCV, with the choice of parameters C, kernel, degree, and gamma. Where C is a multiple of multiplication of 10, gamma multiple 2, and kenel multiple 3, then produce parameters with an accuracy of 81.32% and F1-score of 84.21%..

Keyword: Heart Disease, Support Vector Machine, Logistic Regression, Naïve Bayes, Prediction