

**KLASIFIKASI BAKTERI E. COLI MENGGUNAKAN
ALGORITMA SVM DAN NAÏVE BAYES**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

DAVID NUR PRIHANTORO

21.12.1853

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

KLASIFIKASI BAKTERI E. COLI MENGGUNAKAN ALGORITMA SVM DAN NAÏVE BAYES

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

DAVID NUR PRIHANTORO

21.12.1853

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**KLASIFIKASI BAKTERI E. COLI MENGGUNAKAN ALGORITMA SVM
DAN NAÏVE BAYES**

yang disusun dan diajukan oleh

David Nur Prihantoro

21.12.1853

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 09 Mei 2025

Dosen Pembimbing,


Hanafi, S.Kom, M.Eng, Ph.D
NIK. 190302024

HALAMAN PENGESAHAN

SKRIPSI

**KLASIFIKASI BAKTERI E. COLI MENGGUNAKAN ALGORITMA SVM
DAN NAÏVE BAYES**

yang disusun dan diajukan oleh

David Nur Prihantoro

21.12.1853

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Mei 2025

Susunan Dewan Penguji

Nama Penguji

Windha Mega P. Dhuhita, M.Kom.
NIK. 190302185

Hendra Kurniawan, M.Kom
NIK. 190302244

Hanafi, S.Kom., M.Eng., Ph.D.
NIK. 190302024

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Mei 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : David Nur Prihantoro
NIM : 21.12.1853

Menyatakan bahwa Skripsi dengan judul berikut:

Klasifikasi Bakteri E. Coli Menggunakan Algoritma SVM dan Naïve Bayes

Dosen Pembimbing : Hanafi, S.Kom, M.Eng, Ph.D

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, 20 Mei 2025

Yang Menyatakan,



David Nur Prihantoro

HALAMAN PERSEMBAHAN

Dengan menyebut nama Allah SWT, Yang Maha Pengasih lagi Maha Penyayang, sebagai bentuk rasa syukur, hormat, dan baktiku, skripsi ini saya persembahkan kepada:

1. Kedua orang tua tercinta, Ibu Eny Dwi Siswanti dan Bapak Sumarno, atas segala doa, dukungan, kasih sayang, dan pengorbanan yang tiada henti. Semoga Allah SWT senantiasa melimpahkan kesehatan, rezeki, dan kebahagiaan kepada mereka.
2. Nenek, serta seluruh keluarga besar, khususnya keluarga Pakde Slamet dan Om Agus, yang telah memberikan semangat dan dukungan moral selama proses penyelesaian studi ini.
3. Bapak/Ibu Dosen Universitas AMIKOM Yogyakarta, atas ilmu, bimbingan, dan kesabaran yang telah diberikan selama masa perkuliahan.
4. Bapak M. Hanafi, S.Kom., M.Eng., Ph.D., selaku dosen pembimbing, atas arahan, nasihat, dan koreksi yang sangat berarti dalam proses penulisan skripsi ini.
5. Teman-teman seperjuangan, khususnya Alvin, Faza, Bagas, Jibril dan teman-teman kontrakan lainnya yang tidak dapat saya sebutkan semuanya, yang selalu memberikan semangat, bantuan, dan kebersamaan selama menempuh masa studi hingga selesainya skripsi ini.
6. Semua pihak yang telah memberikan dukungan secara langsung maupun tidak langsung dalam penyelesaian skripsi ini, yang tidak dapat saya sebutkan satu per satu.

Terima kasih atas segala kebaikan dan bantuan yang telah diberikan. Semoga menjadi amal yang diberkahi oleh Allah SWT.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas rahmat dan pertolongan-Nya sehingga skripsi ini dapat diselesaikan dengan baik. Shalawat dan salam tercurah kepada Nabi Muhammad SAW.

Skripsi ini disusun sebagai syarat kelulusan pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta. Penulis mengucapkan terima kasih kepada semua pihak yang telah membantu dan mendukung selama proses penulisan, khususnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak M. Hanafi, S.Kom., M.Eng., Ph.D., selaku dosen pembimbing skripsi yang telah memberikan bimbingan, arahan, dan masukan yang sangat berarti sejak awal hingga terselesaikannya skripsi ini.
3. Ibu Achimah Sidauruk, M.Kom., selaku dosen wali yang telah memberikan nasehat dan pendampingan akademik selama masa studi.
4. Bapak/Ibu Dosen di lingkungan Universitas AMIKOM Yogyakarta yang telah memberikan ilmu, wawasan, dan inspirasi selama penulis menempuh pendidikan.
5. Orang tua tercinta yang selalu memberikan doa, dukungan moral maupun materiil, serta motivasi yang tidak pernah putus.
6. Teman-teman Program Studi Sistem Informasi angkatan 2021, khususnya kelas SI01, atas kebersamaan dan semangat yang saling menguatkan.
7. Seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan bantuan secara langsung maupun tidak langsung dalam penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis membuka diri terhadap saran dan kritik yang membangun demi penyempurnaan di masa mendatang. Semoga karya ini dapat memberikan manfaat bagi pembaca dan menjadi amal kebaikan yang diridhai oleh Allah SWT.

Yogyakarta, 18 Mei 2025

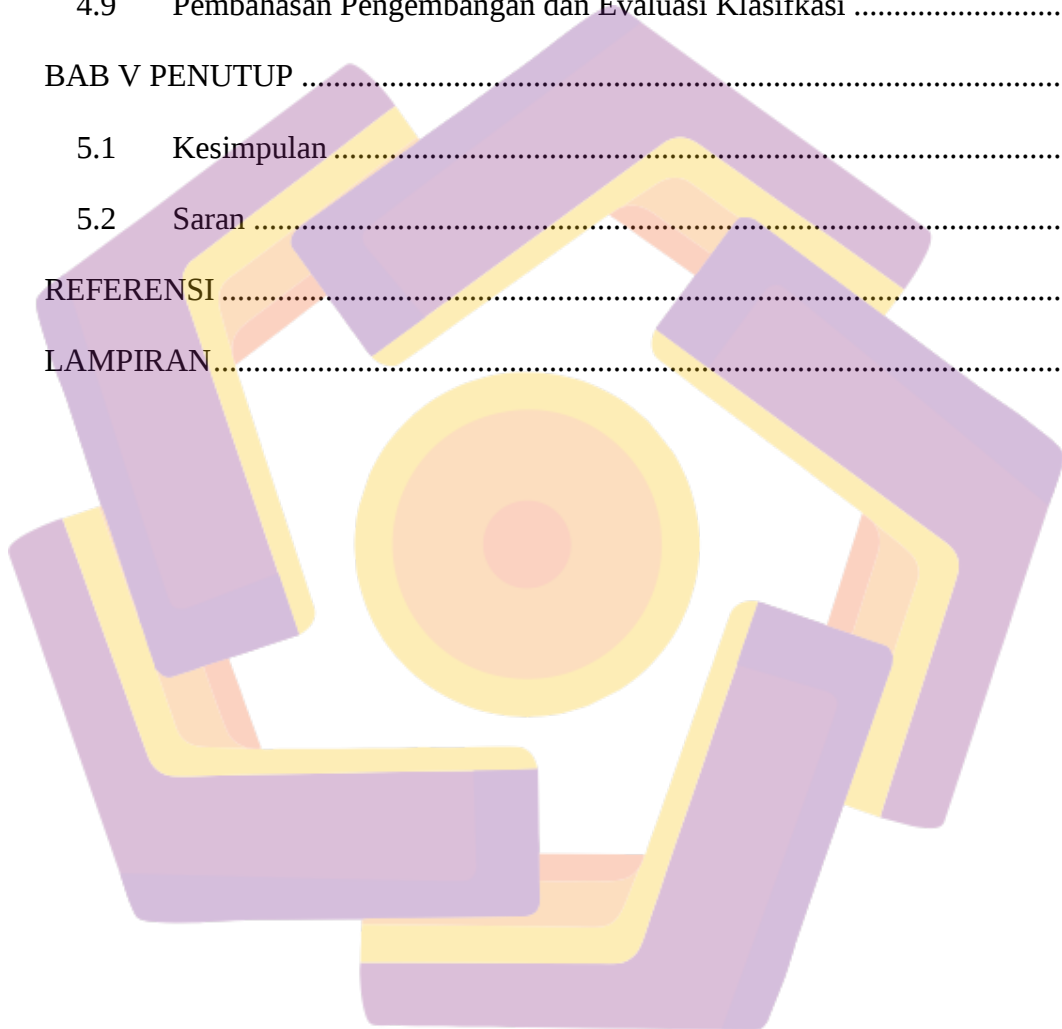
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	Error! Bookmark not defined.
HALAMAN PENGESAHAN	Error! Bookmark not defined.
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	Error! Bookmark not defined.
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	6
1.6 Sistematika Penulisan	7
BAB II TINJAUAN PUSTAKA	9

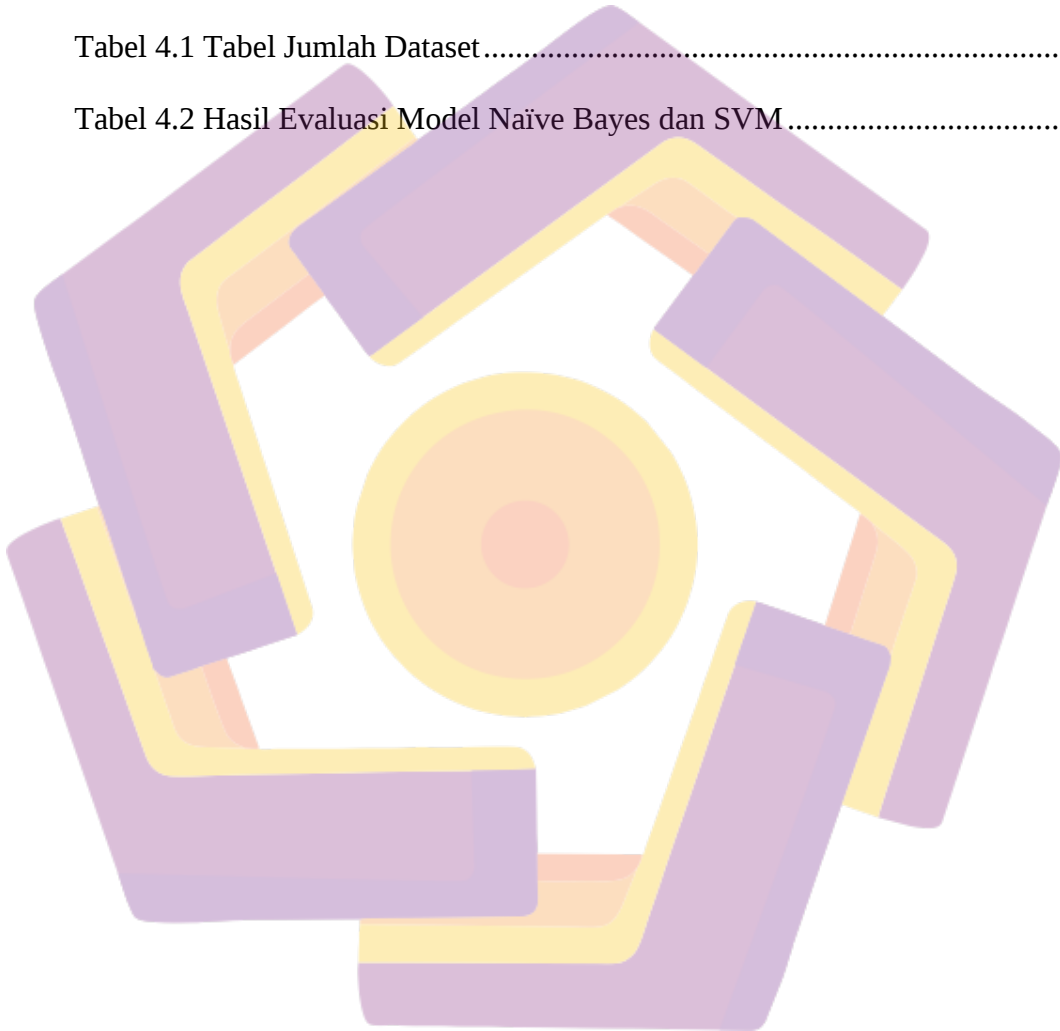
2.1	Studi Literatur	9
2.2	Dasar Teori.....	19
2.2.1	Escherichia coli (E. coli).....	19
2.2.2	Klasifikasi Bakteri dengan Machine Learning	19
2.2.3	Support Vector Machine (SVM).....	19
2.2.4	Naïve Bayes	20
2.2.5	Evaluasi Kinerja Model Machine Learning	20
2.2.6	Python	22
2.2.7	Library.....	24
2.2.8	Dataset.....	26
BAB III METODE PENELITIAN		28
3.1	Alur Penelitian	28
3.2	Pengumpulan Data	30
3.3	Pra-pemrosesan Data	31
3.4	Pengembangan Klasifikasi.....	32
3.5	Evaluasi Klasifikasi	34
3.6	Tools dan Library Pendukung.....	35
3.7	Alat dan Bahan.....	36
3.7.1	Dataset Penelitian.....	36
3.7.2	Software	36
3.7.3	Hardware	36
3.8	Pengembangan Klasifikasi.....	37
BAB IV HASIL DAN PEMBAHASAN		38
4.1	Akses Dataset.....	38
4.2	Deskripsi Dataset	40
4.3	Deskripsi Statistik	42
4.4	Hasil Prapemrosesan Data	47

4.5	Visualisasi Korelasi Fitur.....	52
4.6	Implementasi dan Evaluasi Naive Bayes dalam Klasifikasi.....	56
4.7	Implementasi dan Evaluasi SVM dalam Klasifikasi.....	59
4.8	Perbandingan Performa Sistem Berdasarkan Algoritma	62
4.9	Pembahasan Pengembangan dan Evaluasi Klasifikasi	66
BAB V PENUTUP		71
5.1	Kesimpulan	71
5.2	Saran	73
REFERENSI		74
LAMPIRAN.....		77



DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	13
Tabel 2.2 Confusion Matrix	21
Tabel 2.3 Isi Dataset Lima Baris Pertama.....	27
Tabel 4.1 Tabel Jumlah Dataset	50
Tabel 4.2 Hasil Evaluasi Model Naïve Bayes dan SVM	64

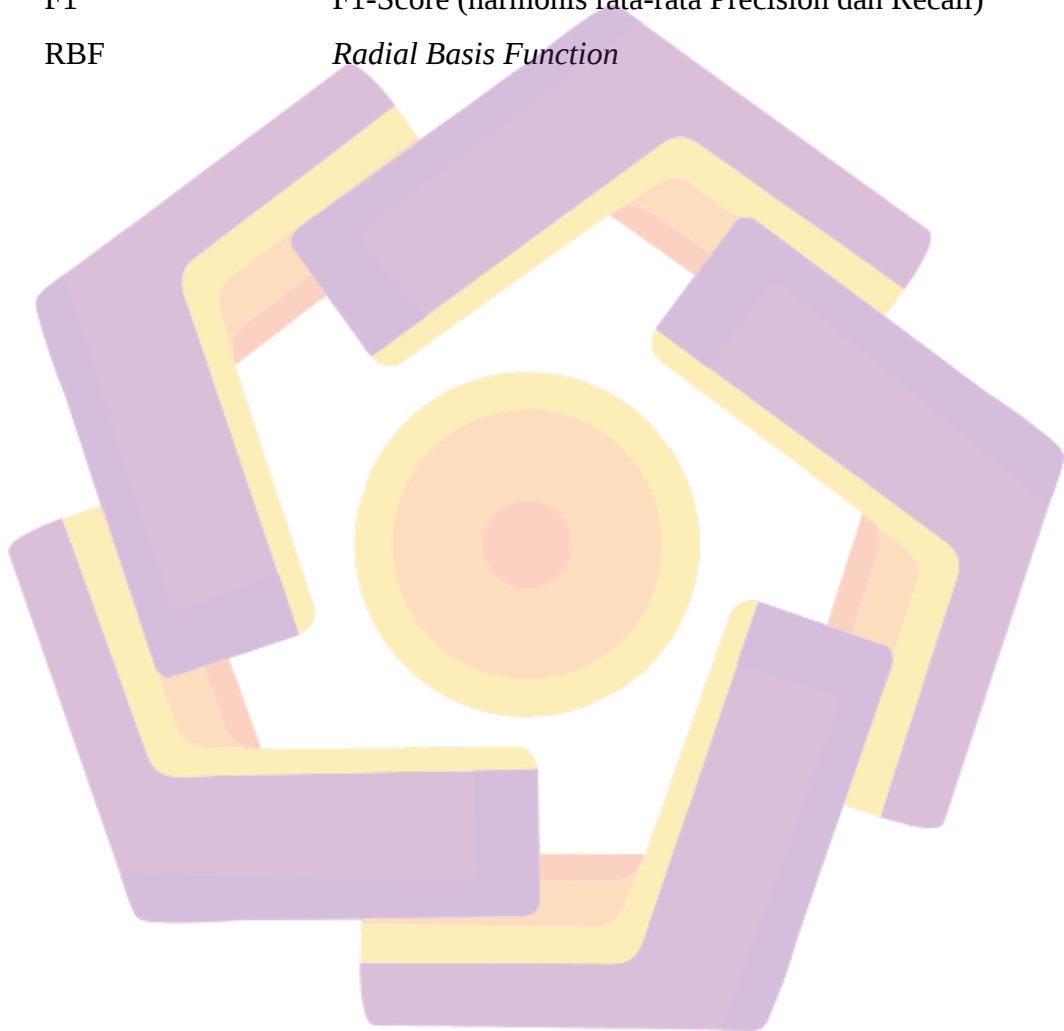


DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	29
Gambar 4.1 Source Code untuk Akses Google Drive	38
Gambar 4.2 Source Code untuk Memanggil dan Menampilkan Dataset.....	39
Gambar 4.3 Hasil Pembacaan dan Penampilan Data.....	40
Gambar 4.4 Hasil untuk Cek Informasi Dataset	42
Gambar 4.5 Hasil untuk Cek Deskripsi Statistik	44
Gambar 4.6 Hasil untuk Cek Missing Value	48
Gambar 4.7 Lima Baris Pertama dari Dataset.....	49
Gambar 4.8 Distribusi Data dalam Bentuk Diagram Batang.....	50
Gambar 4.9 Hasil untuk Visualisasi Korelasi Antar Fitur	54
Gambar 4.10 Confusion Matrix Naïve Bayes.....	57
Gambar 4.11 Confusion Matrix Support Vector Machine.....	60
Gambar 4.12 Hasil untuk Perbandingan Kinerja Naïve Bayes dan SVM	64

DAFTAR LAMBANG DAN SINGKATAN

SVM	<i>Support Vector Machine</i>
NB	Naïve Bayes
CSV	<i>Comma-Separated Values</i>
F1	F1-Score (harmonis rata-rata Precision dan Recall)
RBF	<i>Radial Basis Function</i>



DAFTAR ISTILAH

Klasifikasi	Proses pengelompokan data berdasarkan kategori tertentu.
Evaluasi Model	Proses menilai kinerja model dengan metrik tertentu.
<i>Precision</i>	Proporsi prediksi positif yang benar.
<i>Recall</i>	Proporsi data positif yang berhasil diprediksi benar.
F1 Score	Rata-rata harmonis dari <i>precision</i> dan <i>recall</i> .
Normalisasi	Proses skala ulang data agar berada dalam rentang tertentu.
<i>Feature Extraction</i>	Proses mengambil fitur penting dari data mentah.
Dataset	Kumpulan data yang digunakan untuk pelatihan atau pengujian model.
Akurasi (<i>Accuracy</i>)	Ukuran seberapa tepat model klasifikasi terhadap data aktual.
<i>Confusion Matrix</i>	Matriks yang digunakan untuk mengevaluasi kinerja algoritma klasifikasi.
Algoritma SVM	Metode pembelajaran mesin untuk klasifikasi dengan margin maksimal.
Naïve Bayes	Algoritma klasifikasi berbasis teorema Bayes dengan asumsi independensi.
<i>Deep Learning</i>	Cabang pembelajaran mesin dengan jaringan saraf dalam untuk analisis kompleks.
<i>Overfitting</i>	Model terlalu cocok dengan data latih sehingga buruk pada data baru.
<i>Underfitting</i>	Model terlalu sederhana hingga gagal menangkap pola data.
<i>Reproducible</i>	Hasil dari suatu proses atau eksperimen bisa diulang kembali dan menghasilkan hasil yang sama

INTISARI

Klasifikasi bakteri *Escherichia coli* (E. coli) merupakan langkah penting dalam bidang mikrobiologi dan kesehatan untuk mendeteksi serta menganalisis berbagai strain bakteri yang berpotensi berbahaya. Penelitian ini bertujuan untuk mengembangkan klasifikasi E. coli dengan menggunakan dua algoritma machine learning, yaitu Support Vector Machine (SVM) dan Naïve Bayes. Pengembangan dilakukan melalui penyusunan pipeline klasifikasi yang mencakup preprocessing data, pemisahan data latih dan uji, pelatihan model, evaluasi kinerja, serta visualisasi hasil klasifikasi. Kedua algoritma diterapkan pada dataset E. coli untuk mengevaluasi akurasi, presisi, recall, dan f1-score dalam membedakan strain bakteri berdasarkan karakteristik numeriknya. Hasil eksperimen menunjukkan bahwa SVM memiliki keunggulan dalam menangani data berdimensi tinggi dan memberikan akurasi yang lebih tinggi dibandingkan dengan Naïve Bayes. Namun, Naïve Bayes tetap menunjukkan performa yang kompetitif, terutama dalam kecepatan komputasi dan interpretabilitas. Penelitian ini menegaskan bahwa pengembangan klasifikasi berbasis machine learning dapat memberikan solusi efektif dan efisien dalam identifikasi bakteri *E. coli*, yang berpotensi mendukung penerapan di bidang medis dan mikrobiologi.

Kata kunci: E. coli, klasifikasi, Support Vector Machine, Naïve Bayes, pengembangan.

ABSTRACT

The classification of Escherichia coli (E. coli) bacteria is a crucial step in microbiology and healthcare for identifying and analyzing potentially harmful bacterial strains. This study aims to develop a classification for E. coli using two machine learning algorithms: Support Vector Machine (SVM) and Naïve Bayes. The development process involves building a complete classification pipeline, including data preprocessing, train-test data splitting, model training, performance evaluation, and result visualization. Both algorithms were applied to an E. coli dataset to evaluate their accuracy, precision, recall, and F1-score in distinguishing bacterial strains based on numerical features. The experimental results show that SVM excels in handling high-dimensional data and achieved higher accuracy compared to Naïve Bayes. However, Naïve Bayes remained competitive, particularly in terms of computational speed and model interpretability. This research concludes that a well-developed machine learning-based classification can provide an effective and efficient solution for identifying E. coli bacteria, with potential applications in medical and microbiological fields.

Keywords: *E. coli, classification, Support Vector Machine, Naïve Bayes, development.*