

**KLASIFIKASI BAKTERI E. COLI MENGGUNAKAN
ALGORITMA DECISION TREE DAN RANDOM FOREST**

SKRIPSI

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



disusun oleh

KHOIRUL RIYADI

21.12.1880

Kepada

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

**KLASIFIKASI BAKTERI E. COLI MENGGUNAKAN
ALGORITMA DECISION TREE DAN RANDOM FOREST**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Sistem Informasi



disusun oleh

KHOIRUL RIYADI

21.12.1880

Kepada

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

HALAMAN PERSETUJUAN
SKRIPSI
KLASIFIKASI BAKTERI E. COLI MENGGUNAKAN ALGORITMA DECISION TREE
DAN RANDOM FOREST

yang disusun dan diajukan oleh

Khoirul Riyadi
21.12.1880

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Agustus 2025

Dosen Pembimbing,


Dina Marlina, M. Kom
NIK. 190302250

HALAMAN PENGESAHAN

SKRIPSI

**KLASIFIKASI BAKTERI E. COLI MENGGUNAKAN ALGORITMA DECISION TREE
DAN RANDOM FOREST**

yang disusun dan diajukan oleh

Khoirul Riyadi

21.12.1880

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Agustus 2025

Susunan Dewan Penguji

Nama Penguji

**Irma Rofni Wulandari, S.Pd., M.Eng.
NIK. 190302329**

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

**Dina Maulina, S.Kom., M.Kom.
NIK. 190302250**

Tanda Tangan



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

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Khoirul Riyadi
NIM : 21.12.1880

Menyatakan bahwa Skripsi dengan judul berikut:

Klasifikasi Bakteri E. Coli Menggunakan Algoritma Decision Tree dan Random Forest

Dosen Pembimbing : Dina Maulina, 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, 20 Agustus 2025

Yang Menyatakan,



Khoirul Riyadi

HALAMAN PERSEMBAHAN

Dengan menyebut nama Allah SWT Yang Maha Esa dan Yang Maha Pengasih lagi Maha Penyayang, mengucapkan syukur Alhamdulillah dan segala hormat dan patuh, skripsi ini saya persembahkan kepada:

1. Kedua orang tua tersayang Bapak Suparno dan Ibu Paniyem Purwaningsih yang selalu memberikan kasih sayang, memberikan dukungan moral ataupun materil, dan tak lupa selalu memberikan do'a yang tak pernah putus. Semoga Allah SWT selalu melimpahkan rezeki, kesehatan, dan umur panjang.
2. Keluarga kakak Agus Trisianto, Om Subandi, Om Budi Setiawan dan tak lupa seluruh keluarga nenek dan kakek serta kerabat lainnya yang telah memberikan dukungan semangat selama proses studi.
3. Bu Dina Maulina, M. Kom., telah membimbing saya dari mulai awal hingga akhir skripsi selesai, selalu sabar dan memberikan arahan serta motivasi yang sangat membantu.
4. Bapak/Ibu Dosen yang telah memberikan banyak ilmu dan pengetahuan, tanpa lelah dan selalu sabar untuk memberikan penjelasan sehingga saya dapat memahami materi yang telah disampaikan.
5. Teman seperjuangan, khususnya Sinar, Alvin, David, Faza, Satya, Sandi, dan teman-teman lainnya yang tidak dapat saya sebutkan semuanya. Terima kasih atas dukungan semangat, bantuan, dan kebersamaan selama masa studi dan penyusunan skripsi.

Terima kasih juga kepada semua pihak yang tidak dapat disebutkan satu per satu. Semoga Allah SWT membalaskan kebaikan dan bantuan yang telah diberikan.

KATA PENGANTAR

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

Skripsi ini disusun sebagai salah satu syarat kelulusan pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta. Penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Prof. Dr. Kusri, M. Kom., selaku Dekan Universitas Amikom Yogyakarta.
3. Bapak Anggit Dwi Hartanto, M. Kom., selaku Ketua Program Studi Universitas Amikom Yogyakarta.
4. Ibu Dina Maulina, M. Kom., selaku dosen pembimbing skripsi yang telah memberikan arahan, bimbingan, dan koreksi mulai dari hingga sampai selesai.
5. Bapak Agus Fatkhurohman, M. Kom., selaku dosen wali yang telah mendampingi dan memberikan nasehat terkait akademik selama masa studi.
6. Bapak/Ibu Dosen Universitas Amikom Yogyakarta yang telah memberikan pengetahuan dan wawasan selama masa studi berlangsung.
7. Teman-teman Program Studi Sistem Informasi Khususnya kelas SI01 yang membantu dan memberikan semangat dan kebersamaan selama studi.

Penulis membuka diri terhadap kritik dan saran yang dapat membantu untuk penyempurnaan penelitian selanjutnya dikarenakan skripsi ini jauh dari kata sempurna. Semoga karya ini bermanfaat bagi pembaca dan diradhai Allah SWT.

Yogyakarta, 20 Agustus 2025

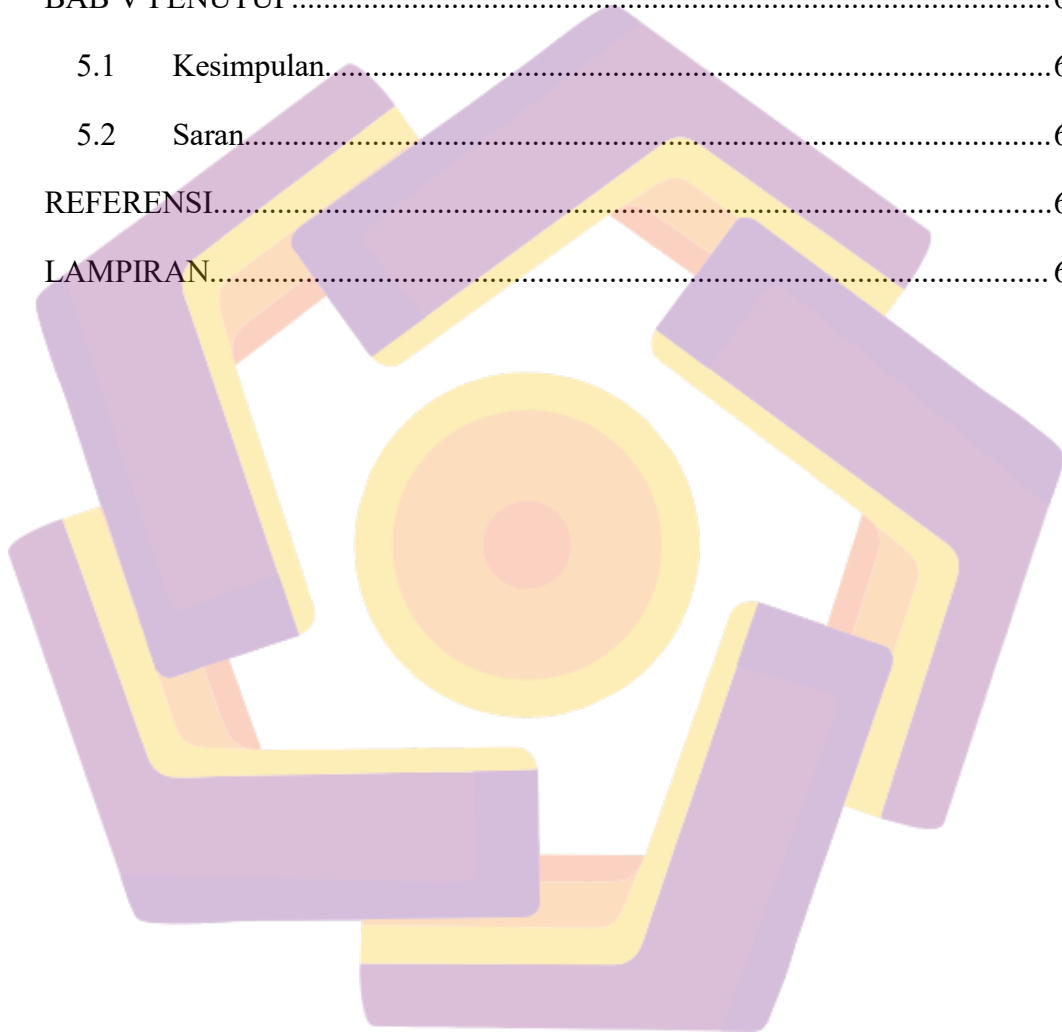
Penulis

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
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.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Studi Literatur.....	6
2.2 Dasar Teori.....	16
2.2.1 Escherichia Coli.....	16
2.2.2 Klasifikasi dalam Machine Learning.....	16

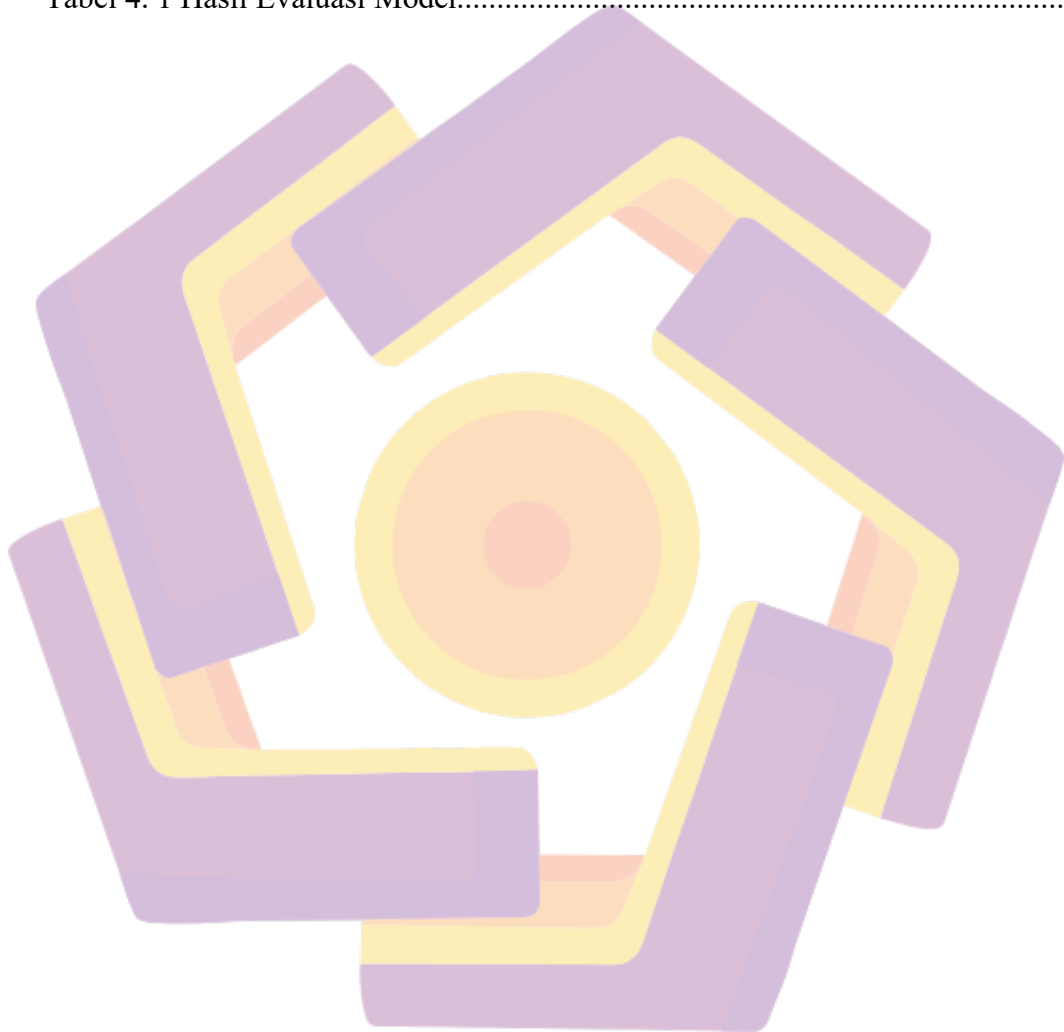
2.2.3 Decision Tree.....	DAFTAR ISI	16
2.2.4 Random Forest.....		17
2.2.5 Evaluasi Kinerja Model <i>Machine Learning</i>		18
2.2.6 Phyton.....		20
2.2.7 Library.....		20
2.2.8 Dataset.....		22
BAB III METODE PENELITIAN.....		23
3.1 Alur Penelitian.....		23
3.2 Pengumpulan Data.....		24
3.3 Pra Pemrosesan Data.....		25
3.4 Evaluasi Pengembangan Klasifikasi.....		26
3.5 Evaluasi Performa Model.....		26
3.6 Analisis Kebutuhan.....		26
3.6.1 Kebutuhan Fungsional.....		26
3.6.2 Kebutuhan Non Fungsional.....		27
BAB IV HASIL DAN PEMBAHASAN.....		28
4.1 Akses Dataset.....		28
4.2 Deskripsi Dataset.....		29
4.2.1 Deskripsi Statistik Dataset.....		32
4.2.2 Korelasi Antar Fitur.....		35
4.3 Pra Pemrosesan Data.....		37
4.4 Evaluasi Klasifikasi Model.....		46
4.4.1 Decision Tree.....		46
4.4.2 Random Forest.....		52
4.5 Evaluasi Performa Model Klasifikasi.....		59
4.6 Kesimpulan Evaluasi Model Klasifikasi.....		62

4.6.1 Akurasi dan <i>FI Score</i>	DAFTAR ISI	62
4.6.2 Waktu Pelatihan Model.....		63
4.6.3 Evaluasi dari <i>Confusion Matrix</i>		63
4.6.4 Analisis Feature Importance.....		63
BAB V PENUTUP.....		64
5.1 Kesimpulan.....		64
5.2 Saran.....		64
REFERENSI.....		65
LAMPIRAN.....		68



DAFTAR TABEL

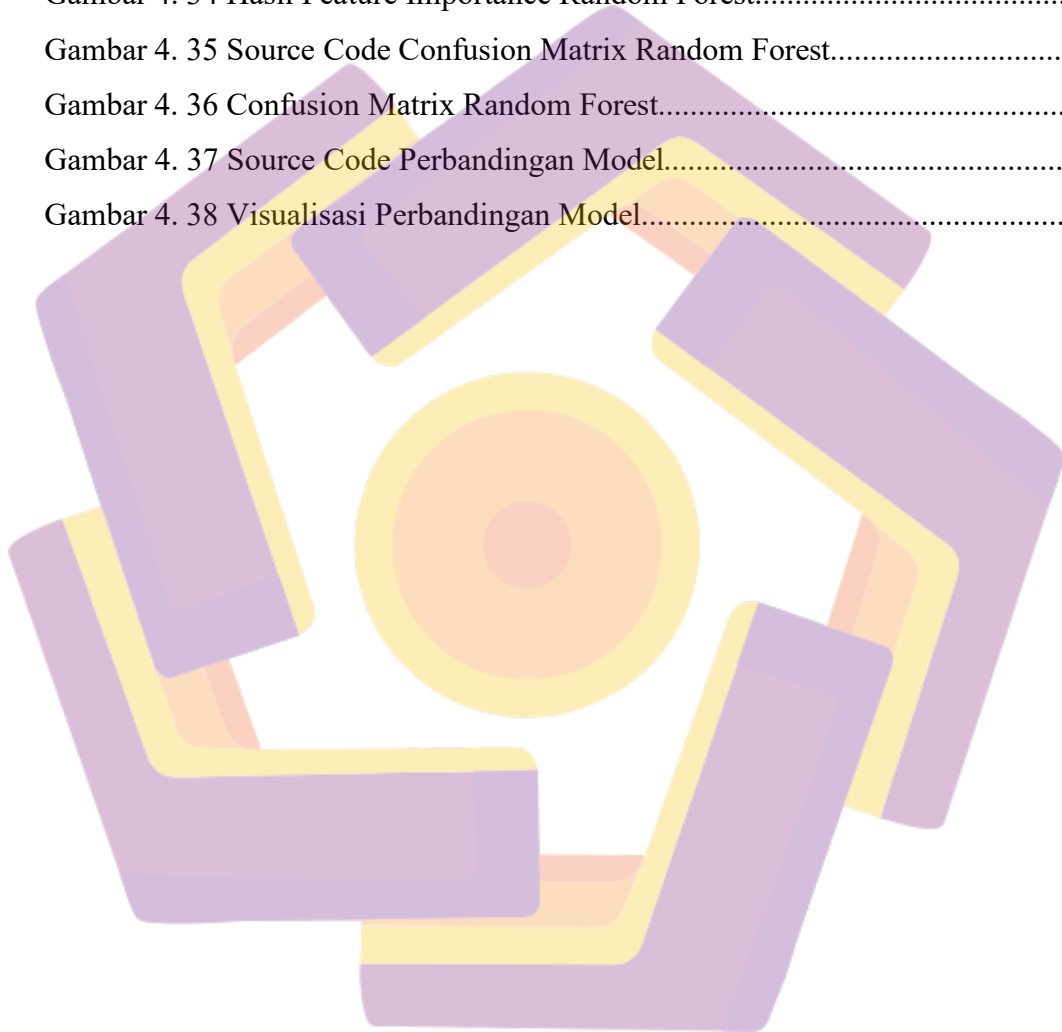
Tabel 2. 1 Keaslian Penelitian.....	10
Tabel 2. 2 Confusin Matrix.....	18
Tabel 2. 3 Isi Dataset Lima Baris Pertama.....	22
Tabel 4. 1 Hasil Evaluasi Model.....	62



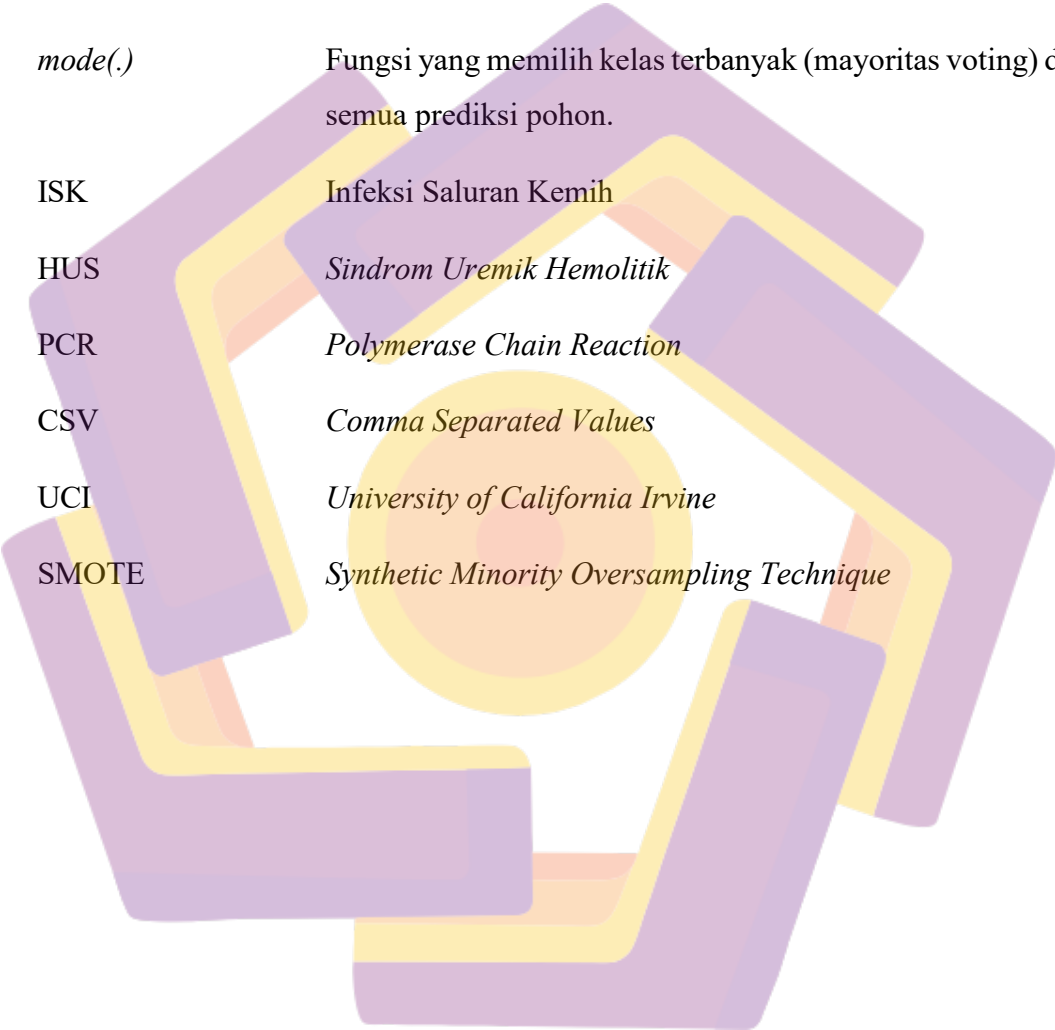
DAFTAR GAMBAR

Gambar 2. 1 Alur Penelitian.....	24
Gambar 4. 1 Kode Untuk Akses Google Drive.....	28
Gambar 4. 2 Google Drive dan Google Colab Terhubung.....	28
Gambar 4. 3 Kode Untuk Memanggil dan Menampilkan Dataset.....	29
Gambar 4. 4 Lima Dataset.....	29
Gambar 4. 5 Source Code Cek Informasi Dataset.....	31
Gambar 4. 6 Informasi Dataset.....	32
Gambar 4. 7 Source Code Cek Deskripsi Statistik.....	33
Gambar 4. 8 Hasil Deskripsi Statistik.....	34
Gambar 4. 9 Source Code Matriks Korelasi Antar Fitur.....	35
Gambar 4. 10 Visualisasi Matriks Korelasi Antar Fitur.....	36
Gambar 4. 11 Source Code Missing Value.....	37
Gambar 4. 12 Hasil Cek Missing Value.....	38
Gambar 4. 13 Source Code Cek Duplikat Data.....	38
Gambar 4. 14 Hasil Cek Duplikat Data.....	38
Gambar 4. 15 Source Code Analisis Distribusi Label.....	39
Gambar 4. 16 Distribusi Data.....	40
Gambar 4. 17 Source Code Normalisasi dan Label Encoding.....	40
Gambar 4. 18 Visualisasi Sebelum dan Sesudah Normalisasi.....	41
Gambar 4. 19 Visualisasi Sebelum dan Sesudah Label Encode.....	42
Gambar 4. 20 Source Code Split Data.....	42
Gambar 4. 21 Visualisasi Split Data.....	43
Gambar 4. 22 Source Code Metode SMOTE.....	44
Gambar 4. 23 Source Code Visualisasi Sebelum dan Sesudah SMOTE.....	44
Gambar 4. 24 Visualisasi Penerapan Metode SMOTE.....	45
Gambar 4. 25 Source Code Klasifikasi Decision Tree.....	46
Gambar 4. 26 Evaluasi Decision Tree.....	47
Gambar 4. 27 Visualisasi Decision Tree.....	48
Gambar 4. 28 Visualisasi Decision Tree.....	49

Gambar 4. 29 Source Code Confusion Matrix Decision Tree.....	50
Gambar 4. 30 Confusion Matrix Decision Tree.....	51
Gambar 4. 31 Source Code Random Forest Classifier.....	53
Gambar 4. 32 Evaluasi Random Forest.....	54
Gambar 4. 33 Source Code Visualisasi Random Forest.....	55
Gambar 4. 34 Hasil Feature Importance Random Forest.....	56
Gambar 4. 35 Source Code Confusion Matrix Random Forest.....	57
Gambar 4. 36 Confusion Matrix Random Forest.....	58
Gambar 4. 37 Source Code Perbandingan Model.....	60
Gambar 4. 38 Visualisasi Perbandingan Model.....	61

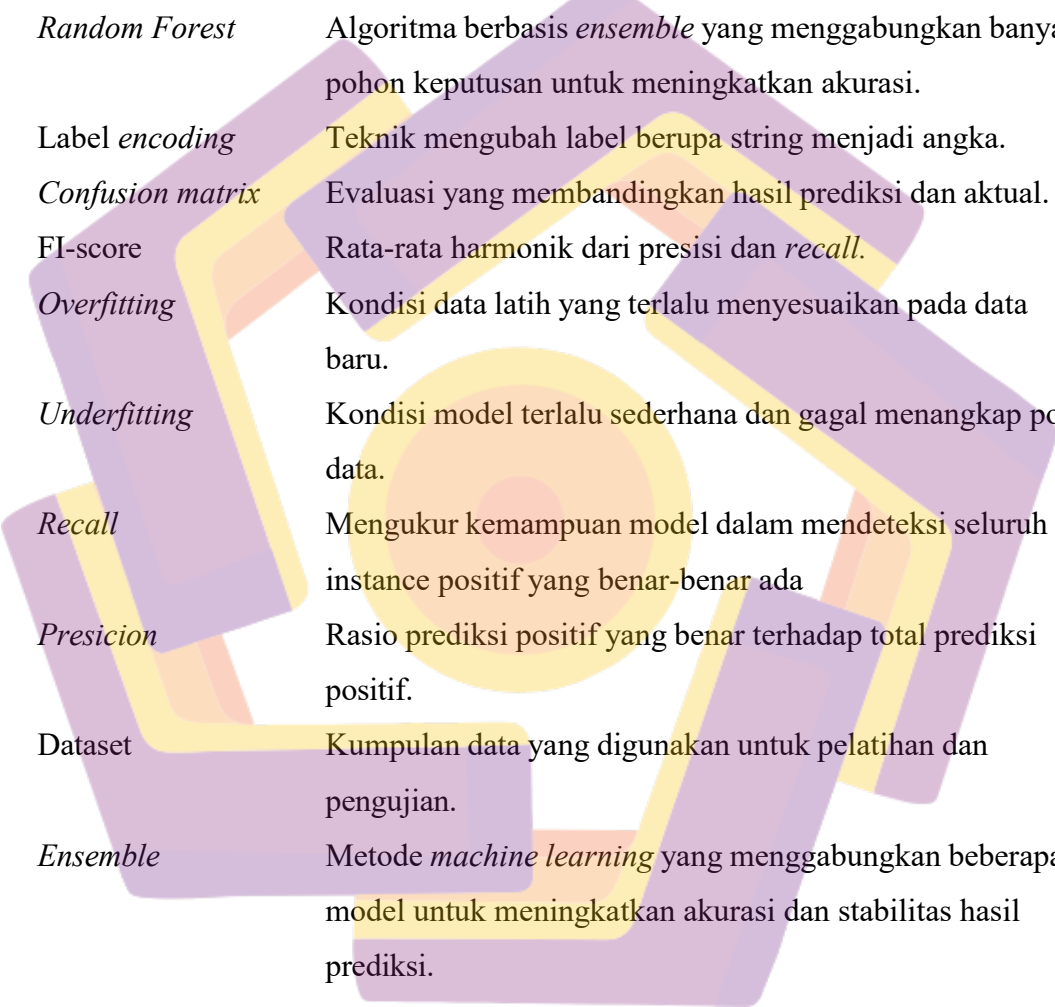


DAFTAR LAMBANG DAN SINGKATAN



y	Hasil prediksi akhir dari <i>random forest</i> .
$h_l(x)$	Prediksi dari pohon keputusan ke- terhadap input .
k	Jumlah total pohon dalam <i>ensemble</i> .
$mode(.)$	Fungsi yang memilih kelas terbanyak (mayoritas voting) dari semua prediksi pohon.
ISK	Infeksi Saluran Kemih
HUS	Sindrom Uremik Hemolitik
PCR	Polymerase Chain Reaction
CSV	Comma Separated Values
UCI	University of California Irvine
SMOTE	Synthetic Minority Oversampling Technique

DAFTAR ISTILAH



<i>Strain</i>	subvarian dari spesies mikroorganisme yang memiliki sifat genetik atau biologis yang khas.
<i>Decision Tree</i>	Algoritma klasifikasi berbentuk pohon yang digunakan untuk mengambil keputusan berdasarkan input data.
<i>Random Forest</i>	Algoritma berbasis <i>ensemble</i> yang menggabungkan banyak pohon keputusan untuk meningkatkan akurasi.
<i>Label encoding</i>	Teknik mengubah label berupa string menjadi angka.
<i>Confusion matrix</i>	Evaluasi yang membandingkan hasil prediksi dan aktual.
<i>F1-score</i>	Rata-rata harmonik dari presisi dan <i>recall</i> .
<i>Overfitting</i>	Kondisi data latih yang terlalu menyesuaikan pada data baru.
<i>Underfitting</i>	Kondisi model terlalu sederhana dan gagal menangkap pola data.
<i>Recall</i>	Mengukur kemampuan model dalam mendeteksi seluruh instance positif yang benar-benar ada
<i>Precision</i>	Rasio prediksi positif yang benar terhadap total prediksi positif.
<i>Dataset</i>	Kumpulan data yang digunakan untuk pelatihan dan pengujian.
<i>Ensemble</i>	Metode <i>machine learning</i> yang menggabungkan beberapa model untuk meningkatkan akurasi dan stabilitas hasil prediksi.
<i>Heterogenitas</i>	Keadaan atau sifat suatu kelompok, campuran, atau sistem yang terdiri dari unsur-unsur yang berbeda atau beragam.
<i>Patogen</i>	Organisme atau mikroorganisme yang dapat menyebabkan penyakit pada organisme lain, termasuk manusia.

INTISARI

Bakteri *Escherichia coli* (*E. coli*) merupakan mikroorganisme yang umum ditemukan dalam usus manusia dan hewan, serta memiliki berbagai *strain* yang bersifat netral maupun patogen. Klasifikasi yang akurat terhadap jenis *E. coli* sangat penting, khususnya dalam bidang kesehatan dan penelitian mikrobiologi. Penelitian ini bertujuan untuk mengklasifikasikan bakteri *E. coli* menggunakan dua algoritma *machine learning*, yaitu *Decision Tree* dan *Random Forest*. Dataset yang digunakan diperoleh dari *UCI Machine Learning Repository*. Model dievaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Hasil pengujian menunjukkan bahwa *Random Forest* memberikan performa lebih baik dengan akurasi sebesar 88.24%, dibandingkan *Decision Tree* yang mencapai 79.41%. Dengan hasil tersebut, *Random Forest* dinilai lebih efektif dan akurat dalam mengklasifikasikan bakteri *E. coli*. Penelitian ini diharapkan dapat membantu untuk dasar pengembangan sistem klasifikasi otomatis dalam bidang biomedis dan diagnostik laboratorium.

Kata kunci: *E. coli*, klasifikasi, *decision tree*, *random forest*

ABSTRACT

Escherichia coli (E. coli) is a microorganism commonly found in the intestines of humans and animals, with various strains that can be either harmless or pathogenic. Accurate classification of E. coli strains is crucial, particularly in the fields of healthcare and microbiological research. This study aims to classify E. coli bacteria using two machine learning algorithms: Decision Tree and Random Forest. The dataset used was obtained from the UCI Machine Learning Repository. The models were evaluated using accuracy, precision, recall, and F1-score metrics. The test results show that Random Forest outperformed Decision Tree, achieving an accuracy of 88.24% compared to 79.41%. Based on these results, Random Forest is considered more effective and accurate in classifying E. coli bacteria. This research is expected to serve as a foundation for the development of automatic classification systems in the biomedical and laboratory diagnostic fields.

Keyword: *E. coli, classification, decision tree, random forest*