

**PENERAPAN METODE K-NEAREST NEIGHBOR UNTUK
KLASIFIKASI STRUKTUR SEL E. COLI BERDASARKAN
DATASET UCI**

SKRIPSI

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



disusun oleh

M. GDE ARY SETIAWAN

21.12.1859

Kepada

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

**PENERAPAN METODE K-NEAREST NEIGHBOR UNTUK
KLASIFIKASI STRUKTUR SEL E. COLI BERDASARKAN
DATASET UCI**

SKRIPSI

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



disusun oleh

M. GDE ARY SETIAWAN

21.12.1859

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**PENERAPAN METODE K-NEAREST NEIGHBOR UNTUK
KLASIFIKASI STRUKTUR SEL E. COLI BERDASARKAN
DATASET UCI**

yang disusun dan diajukan oleh

M. Gde Ary Setiawan

21.12.1859

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 24 November 2025

Dosen Pembimbing,



Bety Wulan Sari, S.Kom., M.Kom
NIK. 190302254

HALAMAN PENGESAHAN
SKRIPSI
PENERAPAN METODE K-NEAREST NEIGHBOR UNTUK
KLASIFIKASI STRUKTUR SEL E. COLI BERDASARKAN
DATASET UCI

yang disusun dan diajukan oleh

M. Gde Ary Setiawan

21.12.1859

Telah dipertahankan di depan Dewan Penguji
pada tanggal 24 November 2025

Susunan Dewan Penguji

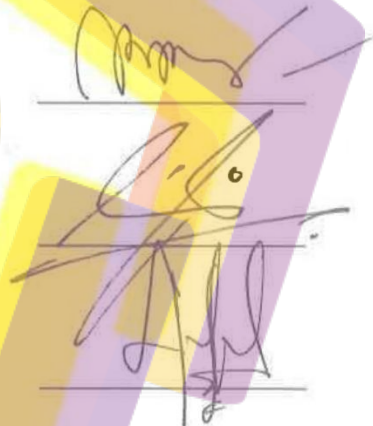
Nama Penguji

Krisnawati, S.Si., M.T.
NIK. 190302038

Agung Nugroho, S.Kom., M.Kom
NIK. 190302242

Bety Wulan Sari, S.Kom., M.Kom
NIK. 190302254

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 24 November 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : M. Gde Ary Setiawan
NIM : 21.12.1859

Menyatakan bahwa Skripsi dengan judul berikut:

Penerapan Metode K-Nearest Neighbor Untuk Klasifikasi Struktur Sel E.Coli Berdasarkan Dataset UCI

Dosen Pembimbing : Bety Wulan Sari, S.Kom., 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, 24 November 2025

Yang Menyatakan,



The image shows a handwritten signature in black ink over a rectangular stamp. The stamp is labeled 'METRAI TEMPEL' and contains the text '02D18ANX207051727'. To the left of the stamp is a vertical strip with the number '10000' and some smaller text.

M. Gde Ary Setiawan

HALAMAN PERSEMBAHAN

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

1. Kedua orang tua tercinta, Bapak Hendra Gunawan dan Ibu Herni Hermiyati, yang selalu memberikan kasih sayang, dukungan moral maupun materi, serta doa yang tak pernah putus. Semoga Allah SWT senantiasa melimpahkan rezeki, kesehatan, dan umur panjang kepada beliau berdua.
2. Ibu Bety Wulan Sari, S.Kom., M.Kom, selaku pembimbing yang telah membimbing saya dari awal hingga skripsi ini selesai dengan penuh kesabaran, memberikan arahan, masukan, serta motivasi yang sangat berarti.
3. Bapak/Ibu Dosen yang telah memberikan ilmu, wawasan, dan bimbingan selama masa studi, dengan kesabaran dan ketulusan sehingga saya dapat memahami setiap materi yang diajarkan.
4. Teman-teman seperjuangan, khususnya Sinar, Irul, Alvin, David, Faza, Satya, Hanif, dan seluruh teman lainnya yang tidak dapat saya sebutkan satu per satu. Terima kasih atas dukungan, kebersamaan, serta bantuan yang telah diberikan selama masa perkuliahan dan penyusunan skripsi ini.

Saya juga menyampaikan terima kasih kepada semua pihak yang telah membantu, mendukung, dan memberikan kebaikan yang tidak dapat disebutkan satu per satu. Semoga Allah SWT membalas setiap kebaikan dengan pahala yang berlipat ganda.

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. Kusrini, M. Kom., selaku Dekan Universitas Amikom Yogyakarta.
3. Bapak Anggit Dwi Hartanto, M. Kom., selaku Ketua Program Studi Universitas Amikom Yogyakarta.
4. Ibu Bety Wulan Sari, S.Kom., M.Kom, selaku dosen pembimbing skripsi yang telah memberikan arahan, bimbingan, dan koreksi mulai dari hingga sampai selesai.
5. Bapak Agung Nugroho, S.Kom., 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, 24 November 2025

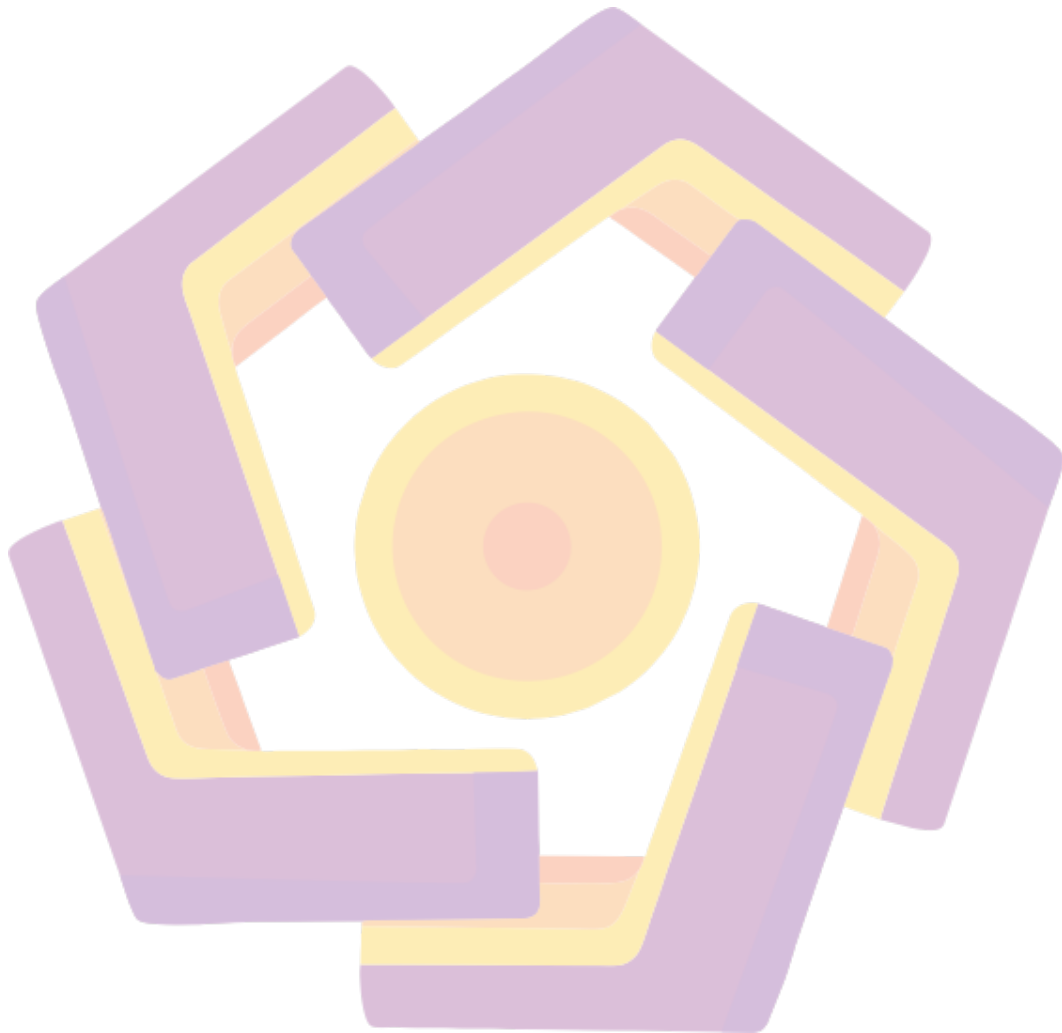
Penulis

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 LAMPIRAN.....	Error! Bookmark not defined.
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xiv
<i>ABSTRACT</i>	xv
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.....	6
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur.....	7

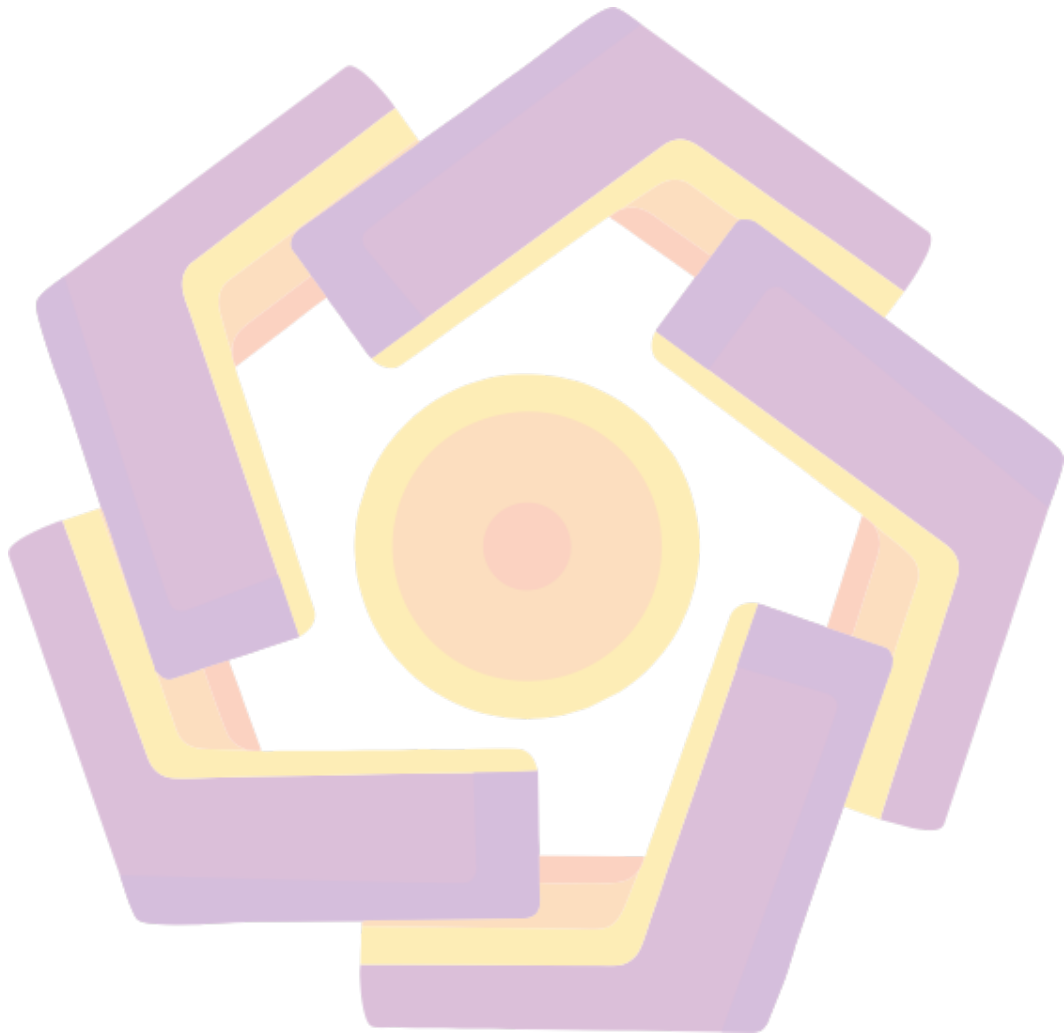
2.2	Dasar Teori	13
2.2.1	Escherichia coli	13
2.2.2	Algoritma K-Nearest Neighbor (K-NN)	13
2.2.3	Pra-Pemrosesan Data dalam K-NN	15
2.2.4	Evaluasi Kinerja K-NN	16
2.2.5	Python.....	17
2.2.6	Library	18
2.2.7	Dataset	19
BAB III METODE PENELITIAN		21
3.1	Alur Penelitian.....	21
3.2	Pengumpulan Data	24
3.3	Pra-Pemrosesan Data.....	25
3.4	Implementasi Algoritma KNN	27
3.5	Evaluasi Model.....	28
3.6	Analisis Hasil KNN.....	29
BAB IV HASIL DAN PEMBAHASAN		31
4.1	Pengumpulan Data	31
4.1.1	Akses Dataset	31
4.1.2	Deskripsi Dataset.....	32
4.1.3	Deskripsi Statistik Dataset.....	36
4.1.4	Korelasi Antar Fitur.....	38
4.2	Pra-Pemrosesan Data.....	41
4.3	Implementasi Algoritma KNN	49
4.3.1	Penentuan Nilai Parameter K	50
4.4	Evaluasi Hasil KNN	54
4.5	Kesimpulan Evaluasi Model KNN	60

BAB V PENUTUP	62
5.1 Kesimpulan.....	62
5.2 Saran.....	62
REFERENSI	64
LAMPIRAN.....	67



DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian	10
Tabel 2.2. <i>Confusion Matrix</i>	15
Tabel 2.3. Isi Dataset LimaBaris Awal	20



DAFTAR GAMBAR

Gambar 3.1. Alur Penelitian	23
Gambar 4.1. Kode Akses Dataset	31
Gambar 4.2. Google Drive dan Google Colab Terhubung	31
Gambar 4.3. Memanggil dan Menampilkan Dataset	32
Gambar 4.4. Lima Dataset	32
Gambar 4.5. <i>Source Code</i> Cek Informasi Dataset	35
Gambar 4.6. Informasi Dataset	35
Gambar 4.7. <i>Source Code</i> Cek Deskripsi Statistik	36
Gambar 4.8. Hasil Deskripsi Statistik	37
Gambar 4.9. <i>Source Code</i> Matriks Korelasi Antar Fitur	39
Gambar 4.10. Visualisasi Matriks Korelasi Antar Fitur	40
Gambar 4.11. <i>Source Code Missing Value</i>	41
Gambar 4.12. Hasil cek <i>Missing Value</i>	41
Gambar 4.13. <i>Source Code</i> Cek Duplikat Data	42
Gambar 4.14. Hasil Cek Duplikasi Data	42
Gambar 4.15. <i>Source Code</i> Analisis Distribusi Label	43
Gambar 4.16. Distribusi Data	44
Gambar 4.17. <i>Source Code</i> Menghapus Outlier(<i>Z-score Method</i>)	44
Gambar 4.18. Hasil Data Sebelum dan Sesudah Pembersihan Outlier	45
Gambar 4.19. <i>Source Code</i> Normalisasi dan Label Encoding	45
Gambar 4.20. Visualisasi Sebelum Normalisasi	46
Gambar 4.21. Visualisasi Sesudah Normalisasi dan Pembersihan Outlier	47
Gambar 4.22. <i>Source Code</i> Data Splitting	48
Gambar 4.23. Visualisasi Data Splitting	48
Gambar 4.24. <i>Source Code</i> Klasifikasi K-Nearest Neighbor	50
Gambar 4.25. Visualisasi Mencari Nilai Parameter K	51
Gambar 4.26. <i>Source Code Confusion Matrix</i> Nilai Parameter K	52
Gambar 4.27. Visualisasi <i>Confusion Matrix</i> Nilai Parameter K	53
Gambar 4.28. <i>Source Code</i> Evaluasi Hasil KNN	54
Gambar 4.29. Visualisasi Evaluasi Nilai Parameter K	55
Gambar 4.30. Visualisasi Performa Model KNN	56
Gambar 4.31. Visualisasi Perbandingan Metrik KNN	57
Gambar 4.32. Visualisasi <i>Heatmap</i> Performa KNN	58

DAFTAR LAMBANG DAN SINGKATAN

x	Titik/data dalam ruang berdimensi.
y	Titik/data dalam ruang berdimensi.
Σ	Penjumlahan dari $i = 1$ sampai n .
K	Jumlah Tetangga Terdekat pada Algoritma K-NN
KNN	Klasifikasikan data berdasarkan tetangga terdekatnya.
UCI	<i>University of California.</i>



DAFTAR ISTILAH

<i>Confusion Matrix</i>	Jumlah prediksi benar dan salah pada tiap kelas.
Dataset	Kumpulan data yang digunakan untuk proses training dan testing model.
<i>Outlier</i>	Nilai yang menyimpang jauh dari pola data normal.
Z-score	Ukuran seberapa jauh suatu nilai menyimpang dari rata-rata dalam satuan standar deviasi.
<i>Accuracy</i>	Metrik untuk menghitung persentase prediksi yang benar dari seluruh data.
<i>Reccal</i>	Kemampuan model data untuk mencari data yang benar-benar positif.
F1-Score	kombinasi antara precision dan recall.

INTISARI

Penelitian ini membahas penerapan metode *K-Nearest Neighbor* (K-NN) untuk melakukan klasifikasi terhadap struktur sel bakteri *Escherichia coli* (E. coli) berdasarkan dataset yang diperoleh dari *University of California, Irvine* (UCI) *Machine Learning Repository*. Tujuan utama dari penelitian ini adalah untuk mengidentifikasi dan mengklasifikasikan jenis struktur sel E. coli berdasarkan sejumlah atribut biologis yang telah tersedia dalam dataset. Metode K-NN dipilih karena kesederhanaannya dalam implementasi serta kemampuannya memberikan hasil klasifikasi yang cukup baik pada data dengan karakteristik yang tidak linier. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan data (normalisasi dan pembagian data), penerapan algoritma K-NN dengan variasi nilai k, serta evaluasi kinerja model menggunakan metrik akurasi, presisi, recall, dan F1-score. Berdasarkan hasil pengujian, metode K-NN menunjukkan tingkat akurasi yang cukup tinggi dalam mengklasifikasikan struktur sel E. coli, dengan nilai akurasi terbaik diperoleh pada k tertentu yang menghasilkan keseimbangan optimal antara bias dan variansi. Hasil penelitian ini menunjukkan bahwa metode K-NN dapat digunakan secara efektif dalam menganalisis dan mengklasifikasikan data biologis, khususnya dalam konteks struktur sel E. coli. Temuan ini diharapkan dapat menjadi acuan dalam pengembangan sistem klasifikasi berbasis machine learning di bidang biologi mikroorganisme.

Kata kunci: *K-Nearest Neighbor*, Klasifikasi, E. coli, *Dataset UCI*, *Machine Learning*

ABSTRACT

This study discusses the application of the K-Nearest Neighbor (K-NN) method to classify the cellular structures of Escherichia coli (E. coli) bacteria based on a dataset obtained from the University of California, Irvine (UCI) Machine Learning Repository. The main objective of this research is to identify and classify the types of E. coli cell structures using a set of biological attributes provided in the dataset. The K-NN method was chosen due to its simplicity of implementation and its ability to produce reliable classification results on data with non-linear characteristics. The research stages include data collection, data preprocessing (normalization and data splitting), application of the K-NN algorithm with varying k values, and model performance evaluation using accuracy, precision, recall, and F1-score metrics. Based on the experimental results, the K-NN method achieved a relatively high accuracy in classifying E. coli cell structures, with the best performance obtained at an optimal k value that balances bias and variance. The results of this study indicate that the K-NN method can be effectively used to analyze and classify biological data, particularly in the context of E. coli cell structure classification. These findings are expected to serve as a reference for developing machine learning-based classification systems in the field of microorganism biology.

Keyword: K-Nearest Neighbor, Classification, E. coli, UCI Dataset, Machine Learning