

**PERBANDINGAN PERFORMA METODE KLASIFIKASI
BERBASIS *MACHINE LEARNING* DALAM PENENTUAN
TINGKAT PENCEMARAN AIR**

SKRIPSI

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



disusun oleh

WAHYU MUHAMMAD ARBY

22.12.2553

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**PERBANDINGAN PERFORMA METODE KLASIFIKASI
BERBASIS *MACHINE LEARNING* DALAM PENENTUAN
TINGKAT PENCEMARAN AIR**

SKRIPSI

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



disusun oleh

WAHYU MUHAMMAD ARBY

22.12.2553

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN PERFORMA METODE KLASIFIKASI BERBASIS
MACHINE LEARNING DALAM PENENTUAN TINGKAT
PENCEMARAN AIR**

yang disusun dan diajukan oleh

Wahyu Muhammad Arby

22.12.2553

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Februari 2026

Dosen Pembimbing,



Afrig Aminuddin, S.Kom., M.Eng., Ph.D.

NIK. 190302351

HALAMAN PENGESAHAN

SKRIPSI

**PERBANDINGAN PERFORMA METODE KLASIFIKASI BERBASIS
MACHINE LEARNING DALAM PENENTUAN TINGKAT
PENCEMARAN AIR**

yang disusun dan diajukan oleh

Wahyu Muhammad Arby

22.12.2553

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Februari 2026

Susunan Dewan Penguji

Nama Penguji

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

Alfie Nur Rahmi, S.Kom., M.Kom.
NIK. 190302240

Afrig Aminuddin, S.Kom., M.Eng., Ph.D.
NIK. 190302351

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Wahyu Muhammad Arby
NIM : 22.12.2553

Menyatakan bahwa Skripsi dengan judul berikut:

Perbandingan Performa Metode Klasifikasi Berbasis *Machine learning* dalam Penentuan Tingkat Pencemaran Air

Dosen Pembimbing : Afrig Aminuddin, 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, 19 Februari 2026

Yang Menyatakan,



Wahyu Muhammad Arby

HALAMAN PERSEMBAHAN

Karya tulis ini penulis persembahkan sebagai wujud rasa bakti, hormat, dan cinta kasih yang tulus kepada:

1. Kedua orang tua tercinta, Bapak Drs. Muhamad Tamrun dan Ibu Harneti, yang senantiasa melantunkan doa di setiap sujudnya, serta memberikan pengorbanan, kasih sayang, dan dukungan luar biasa yang menjadi sumber kekuatan terbesar bagi penulis dalam menapaki setiap langkah kehidupan dan pendidikan.
2. Saudara-saudara kandung saya, Hanum Mutmainah, Fadel Fatur Rahman dan Muhammad Rafat Kafkah Fiqam yang selalu memberikan semangat, keceriaan, dan dukungan moral sehingga penulis tetap tegar dalam menyelesaikan studi ini.
3. Teman-teman seperjuangan, khususnya kelas SI-05 angkatan 22 yang telah menjadi kawan diskusi, memberikan bantuan teknis, serta berbagi suka dan duka dalam proses penyelesaian skripsi ini. Terima kasih atas kebersamaan dan persahabatan yang tak ternilai harganya.
4. Grup idol K-Pop Red Velvet (Irene, Seulgi, Wendy, Joy, dan Yeri). Terima kasih atas ragam karya musik dan konten inspiratif yang telah menjadi teman setia, penglipur lara di saat penat, serta pembangkit semangat tersendiri bagi penulis untuk kembali bangkit dan menyelesaikan penyusunan skripsi ini.

Semoga persembahan sederhana ini dapat menjadi bukti kecil atas besarnya rasa terima kasih penulis kepada pihak-pihak yang sangat berarti dalam perjalanan hidup penulis. Semoga Tuhan Yang Maha Esa senantiasa melimpahkan rahmat dan kebaikan-Nya kepada kita semua.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena atas limpahan rahmat, hidayah, dan karunia-Nya, penulis dapat menyelesaikan penyusunan skripsi yang berjudul “Perbandingan Performa Metode Klasifikasi Berbasis *Machine learning* dalam Penentuan Tingkat Pencemaran Air” tepat pada waktunya.

1. Bapak Prof. Dr. Mohammad Suyanto, M.M., selaku Rektor Universitas AMIKOM Yogyakarta, yang telah memberikan kesempatan kepada penulis untuk menimba ilmu di perguruan tinggi ini.
2. Bapak Afrig Aminuddin, S.Kom., M.Eng., Ph.D, selaku Dosen Pembimbing. Terima kasih yang tak terhingga penulis sampaikan atas kesabaran, waktu, tenaga, dan pikiran yang telah diluangkan untuk memberikan bimbingan, koreksi, serta motivasi yang sangat berharga sejak awal perencanaan hingga terselesaikannya penulisan skripsi ini.
3. Teristimewa kepada kedua orang tua dan keluarga besar penulis. Tidak ada kata yang cukup untuk melukiskan rasa terima kasih penulis atas segala doa restu, kasih sayang, serta dukungan materiil maupun moril yang diberikan tanpa henti.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan jauh dari kesempurnaan, baik dari segi materi maupun tata bahasa. Oleh karena itu, kritik dan saran yang bersifat membangun sangat penulis harapkan demi perbaikan di masa yang akan datang. Akhir kata, semoga skripsi ini dapat memberikan manfaat dan kontribusi positif bagi pengembangan ilmu pengetahuan serta bagi pihak-pihak yang membutuhkan.

Yogyakarta, 19 Februari 2026

Penulis

DAFTAR ISI

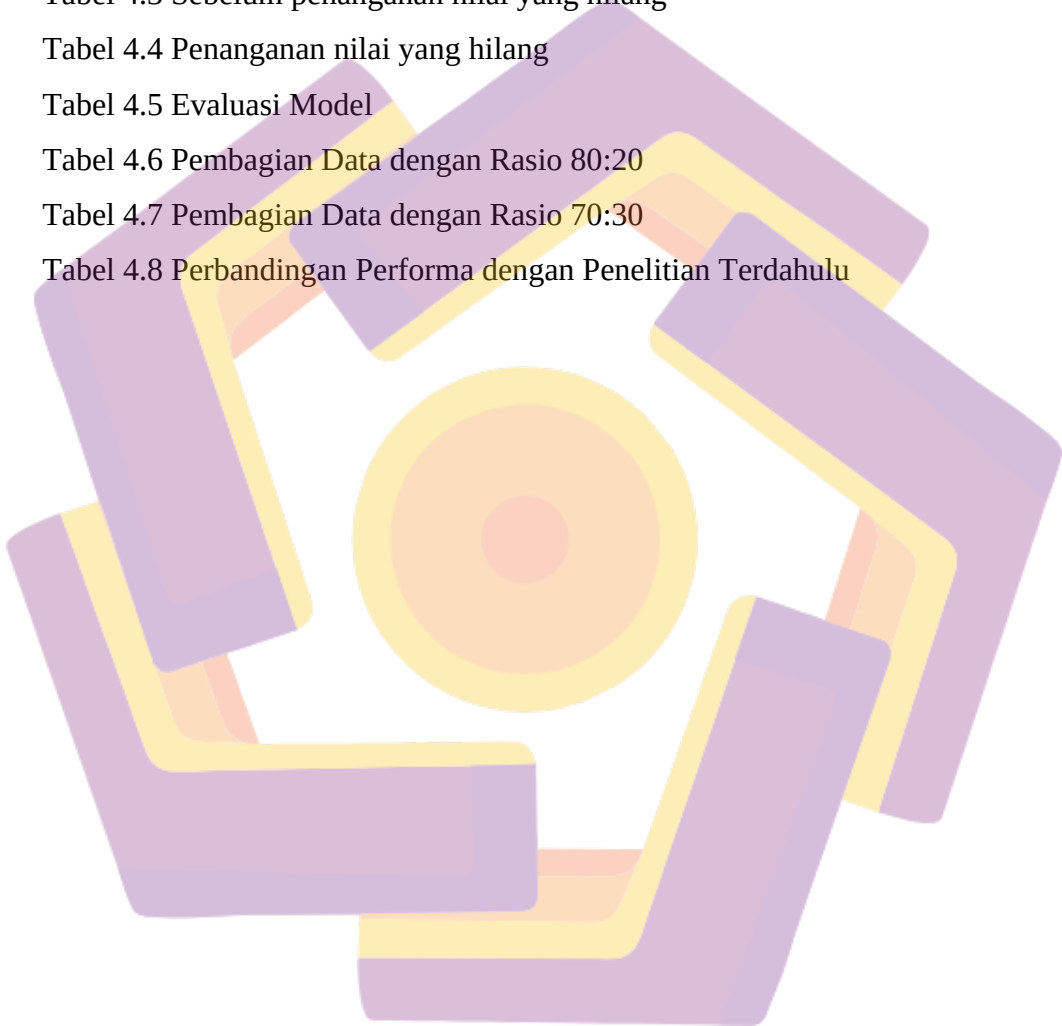
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN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

BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	11
2.2.1 Distribusi Fitur dengan Kelas Potabilitas	11
2.2.2 <i>Machine learning</i> dan Klasifikasi	12
2.2.3 Pra-pemrosesan Data (Data Pre-processing)	13
2.2.4 Algoritma Klasifikasi	17
2.2.5 Evaluasi Kinerja	19
BAB III METODE PENELITIAN	21
3.1 Objek Penelitian	21
3.2 Alur Penelitian	22
3.2.1 Pengumpulan Data	23
3.2.2 Eksplorasi Data (EDA)	24
3.2.3 Pra-pemrosesan Data	24
3.2.4 Analisis Model Training	28
3.2.5 Evaluasi Performa Model (<i>Performance Evaluation</i>)	28
3.3 Alat dan Bahan	29
BAB IV HASIL DAN PEMBAHASAN	30
4.1 Pengumpulan Data	30
4.2 Eksplorasi Data (EDA)	30
4.2.1 Struktur Data	31
4.2.2 Statistik Deskriptif	32

4.2.3	Distribusi Variabel Dependent	33
4.2.4	Distribusi Fitur dengan Kelas Potabilitas	34
4.2.5	Korelasi Antar Fitur	36
4.3	Pra-pemrosesan Data	38
4.3.1	Penanganan Nilai yang Hilang (<i>Missing Values</i>)	39
4.3.2	Penanganan Ketidakseimbangan Data (<i>Resamples</i>)	42
4.3.3	Normalisasi Data	45
4.3.4	Seleksi Fitur	47
4.4	Analisis Model Training	49
4.5	Evaluasi Performa Model (<i>Perfomance Evaluation</i>)	52
4.6	Analisis Kurva ROC dan <i>Confusion Matrix</i>	60
4.6.1	Kurva ROC	60
4.6.2	<i>Confusion Matrix</i>	62
4.7	Deploy	64
4.7.1	Halaman Beranda	64
4.7.2	Halaman Prediksi Pencemaran Air	66
4.8	Pembahasan	69
BAB V PENUTUP		74
5.1	Kesimpulan	74
5.2	Saran	75
REFERENSI		76
LAMPIRAN		80

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	8
Tabel 4.1 Struktur dan tipe data dataset	31
Tabel 4.2 Statistik deskriptif	32
Tabel 4.3 Sebelum penanganan nilai yang hilang	40
Tabel 4.4 Penanganan nilai yang hilang	41
Tabel 4.5 Evaluasi Model	53
Tabel 4.6 Pembagian Data dengan Rasio 80:20	56
Tabel 4.7 Pembagian Data dengan Rasio 70:30	57
Tabel 4.8 Perbandingan Performa dengan Penelitian Terdahulu	71

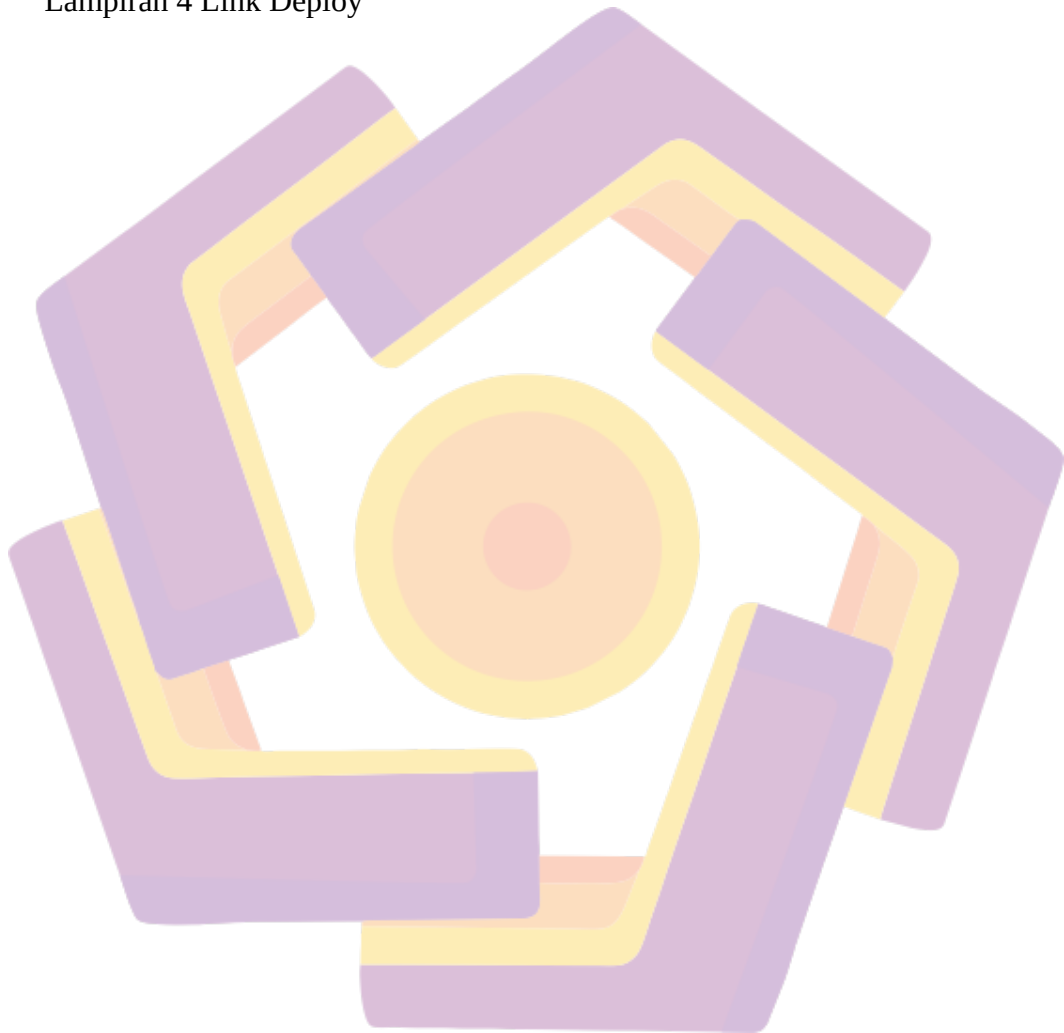


DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	22
Gambar 4.1 Diagram lingkaran fitur kelayakan minum	33
Gambar 4.2 Distribusi densitas fitur berdasarkan kelas potabilitas	35
Gambar 4.3 Matriks Korelasi Pearson Antar Fitur Kualitas Air	37
Gambar 4.4 Distribusi kelas setelah <i>Random oversampling</i>	44
Gambar 4.5 Hasil normalisasi data	46
Gambar 4.6 Hasil seleksi fitur	48
Gambar 4.7 Source code pemodelan	50
Gambar 4.8 Metriks evaluasi akurasi dan <i>F1-Score</i>	59
Gambar 4.9 Kurva ROC seluruh model klasifikasi	61
Gambar 4.10 <i>Confusion Matrix</i> seluruh model klasifikasi	63
Gambar 4.11 Tampilan halaman beranda aplikasi	65
Gambar 4.12 Hasil prediksi kategori layak minum	67
Gambar 4.13 Hasil prediksi kategori tidak layak minum	68

DAFTAR LAMPIRAN

Lampiran 1. Dataset	80
Lampiran 2. Source Code	80
Lampiran 3 Source Code Streamlit	81
Lampiran 4 Link Deploy	81

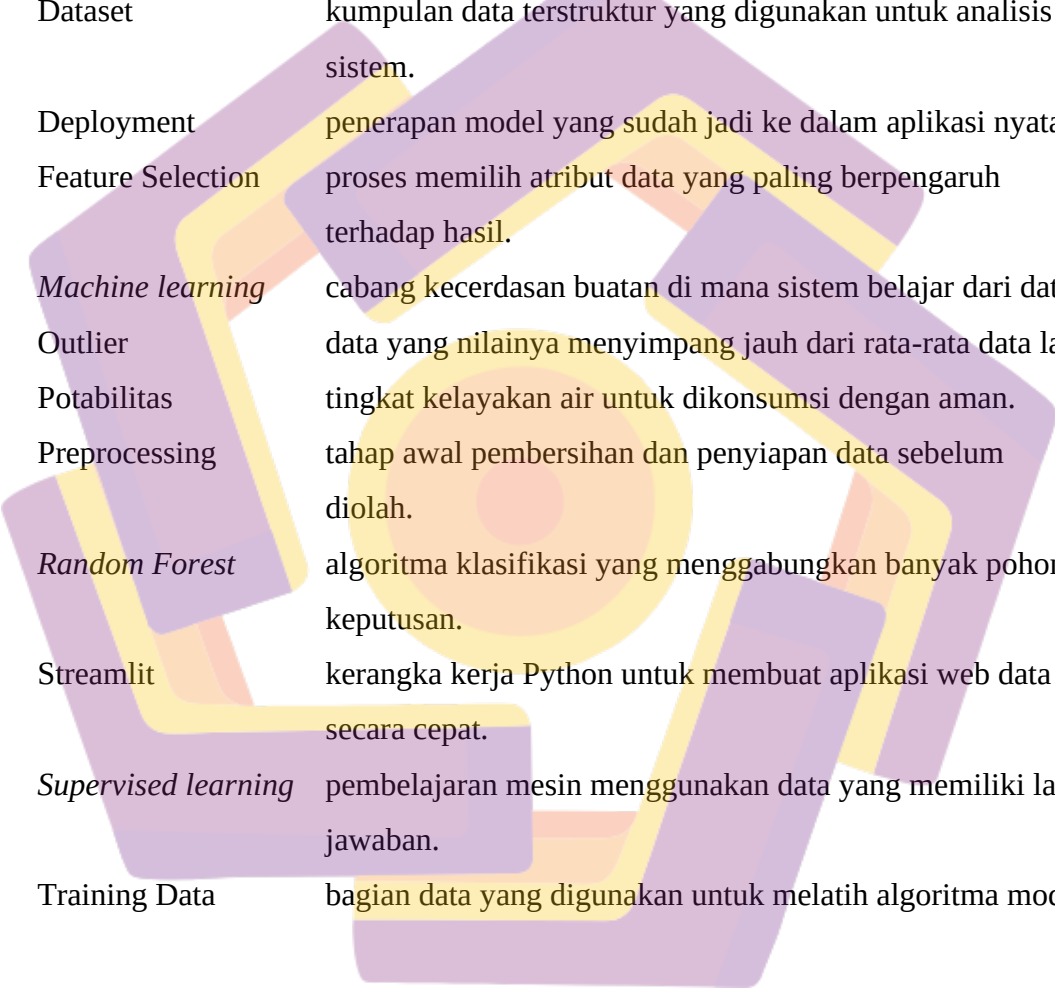


DAFTAR LAMBANG DAN SINGKATAN



ANOVA	Analysis of Variance
AUC	<i>Area Under Curve</i>
EDA	Exploratory Data Analysis
FN	False Negative
FP	False Positive
F1	<i>F1-Score</i> (Rata-rata harmonis Presisi dan <i>Recall</i>)
KDD	Knowledge Discovery in Databases
KNN	<i>K-Nearest Neighbors</i>
mg/L	Milligram per liter (satuan konsentrasi)
μ	<i>Mean</i> (Rata-rata populasi)
NTU	<i>NepHelometric Turbidity Unit</i> (satuan kekeruhan)
n	Jumlah sampel data
pH	<i>Potential of Hydrogen</i> (Derajat Keasaman)
ppm	<i>Parts per million</i> (bagian per juta)
ROC	Receiver Operating Characteristic
σ	Standar Deviasi
SVM	<i>Support Vector Machine</i>
TDS	Total Dissolved Solid
THMs	<i>Trihalomethanes</i>
TN	True Negative
TP	True Positive
WHO	World Health Organization
<i>XGBoost</i>	Extreme Gradient Boosting
$\mu S/cm$	Microsiemens per centimeter (satuan konduktivitas)

DAFTAR ISTILAH



<i>Accuracy</i>	ketepatan model dalam memprediksi data secara benar.
<i>Confusion Matrix</i>	tabel evaluasi untuk membandingkan hasil prediksi dengan data aktual.
Data Mining	proses penggalian pola penting dari sekumpulan data besar.
Dataset	kumpulan data terstruktur yang digunakan untuk analisis sistem.
Deployment	penerapan model yang sudah jadi ke dalam aplikasi nyata.
Feature Selection	proses memilih atribut data yang paling berpengaruh terhadap hasil.
<i>Machine learning</i>	cabang kecerdasan buatan di mana sistem belajar dari data.
Outlier	data yang nilainya menyimpang jauh dari rata-rata data lain.
Potabilitas	tingkat kelayakan air untuk dikonsumsi dengan aman.
Preprocessing	tahap awal pembersihan dan penyiapan data sebelum diolah.
<i>Random Forest</i>	algoritma klasifikasi yang menggabungkan banyak pohon keputusan.
Streamlit	kerangka kerja Python untuk membuat aplikasi web data secara cepat.
<i>Supervised learning</i>	pembelajaran mesin menggunakan data yang memiliki label jawaban.
Training Data	bagian data yang digunakan untuk melatih algoritma model.

INTISARI

Penilaian pencemaran air merupakan tugas yang sangat penting dalam pemantauan lingkungan karena dampaknya yang langsung terhadap kesehatan manusia dan keberlanjutan ekosistem. Penelitian ini menganalisis kinerja metode klasifikasi berbasis *machine learning* dalam menentukan tingkat pencemaran air menggunakan data kualitas air yang tersedia secara publik. Beberapa algoritma *supervised learning*, antara lain *Decision Tree*, *Random Forest*, *Support Vector Machine*, *K-Nearest Neighbor*, dan *Naïve Bayes*, diterapkan untuk mengklasifikasikan tingkat pencemaran berdasarkan parameter fisikokimia air. Dataset dievaluasi menggunakan skema pembagian data latih dan data uji dengan rasio 90:10, serta kinerja model dinilai menggunakan metrik *accuracy*, *precision*, *Recall*, *F1-Score*, dan *ROC-AUC*. Hasil eksperimen menunjukkan bahwa pengklasifikasi berbasis ensemble menghasilkan kinerja klasifikasi yang lebih unggul dibandingkan pengklasifikasi tunggal, terutama dalam menangkap hubungan nonlinier antar indikator kualitas air. Temuan ini sejalan dengan penelitian-penelitian terkini yang menekankan efektivitas teknik *machine learning* dalam klasifikasi kualitas air dan penilaian pencemaran air.

Kata kunci: pencemaran air, *machine learning*, klasifikasi, analisis performa, kualitas air.

ABSTRACT

Water pollution assessment is a critical task in environmental monitoring due to its direct impact on human health and ecosystem sustainability. This study analyzes the performance of machine learning–based classification methods for determining Water pollution levels using publicly available Water Quality data. Several supervised learning algorithms, including Decision Tree, Random Forest, Support Vector Machine , K-Nearest Neighbor, and Naïve Bayes, were implemented to classify pollution levels based on pHysicochemical Water parameters. The dataset was evaluated using a 90:10 train–test split, and model performance was assessed using accuracy, precision, Recall, F1-Score, and ROC-AUC metrics. The experimental results indicate that ensemble-based classifiers achieve superior classification performance compared to single classifiers, particularly in capturing nonlinear relationships among Water Quality indicators. These findings are consistent with recent studies that empHasize the effectiveness of machine learning techniques for Water Quality classification and pollution assessment.

Keyword: Water pollution, machine learning, classification, performance analysis, Water Quality.