

**OPTIMASI METODE ENSEMBLE STACKING DALAM
KLASIFIKASI KUALITAS AIR BERBASIS ALGORITMA
XGBOOST, LIGHTGBM, DAN CATBOOST**

SKRIPSI

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



disusun oleh

RAIHAN ALY ZAKY

22.11.4667

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**OPTIMASI METODE ENSEMBLE STACKING DALAM
KLASIFIKASI KUALITAS AIR BERBASIS ALGORITMA
XGBOOST, LIGHTGBM, DAN CATBOOST**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

RAIHAN ALY ZAKY

22.11.4667

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**OPTIMASI METODE ENSEMBLE STACKING DALAM KLASIFIKASI
KUALITAS AIR BERBASIS ALGORITMA XGBOOST, LIGHTGBM,
DAN CATBOOST**

yang disusun dan diajukan oleh

Raihan Aly Zaky

22.11.4667

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 27 Januari 2026

Dosen Pembimbing,



Windha Mega Pradnya Dhuhiita, S.Kom., M.Kom.
NIK. 190302185

HALAMAN PENGESAHAN
SKRIPSI
OPTIMASI METODE ENSEMBLE STACKING DALAM KLASIFIKASI
KUALITAS AIR BERBASIS ALGORITMA XGBOOST, LIGHTGBM,
DAN CATBOOST

yang disusun dan diajukan oleh

Raihan Aly Zaky

22.11.4667

Telah dipertabankan di depan Dewan Penguji
pada tanggal 27 Januari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dr. Hartatik, S.T., M.Cs.
NIK. 190302232



Bambang Pili Hartato, S.Kom., M.Eng.
NIK. 190302707



Windha Mega Pradnya Dhubita, S.Kom., M.Kom.
NIK. 190302185



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 27 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Raihan Aly Zaky
NIM : 22.11.4667

Menyatakan bahwa Skripsi dengan judul berikut:

OPTIMASI METODE ENSEMBLE STACKING DALAM KLASIFIKASI KUALITAS AIR BERBASIS ALGORITMA XGBOOST, LIGHTGBM, DAN CATBOOST

Dosen Pembimbing : Nama Dosen dan Gelar

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, 27 Januari 2026

Yang Menyatakan,



Raihan Aly Zaky

HALAMAN PERSEMBAHAN

Dengan menyebut nama Allah Subhanahu wa Ta'ala Yang Maha Pengasih lagi Maha Penyayang, segala puji dan syukur senantiasa tercurah atas segala rahmat dan karunia yang telah dilimpahkan, sehingga Skripsi ini dapat diselesaikan dengan baik.

Karya ini dengan segala kerendahan hati dan rasa syukur yang mendalam, penulis persembahkan kepada:

1. Allah Subhanahu wa Ta'ala, atas segala petunjuk, kekuatan, dan kemudahan yang tak terhingga dalam setiap langkah penulisan skripsi ini.
2. Kedua Orang Tua tercinta, Abah dan Ibu, yang telah melimpahkan kasih sayang, doa tulus tanpa batas, dukungan moral dan material yang tak pernah putus, serta menjadi sumber inspirasi dan motivasi terbesar dalam hidup penulis.
3. Keluarga Besar, yang selalu memberikan semangat, tawa, dan dukungan tak ternilai
4. selama proses penyelesaian studi.
5. Almamater tercinta, Universitas AMIKOM Yogyakarta.

Penelitian ini diharapkan dapat membawa manfaat, memperluas pengetahuan, dan menjadi langkah awal dalam mendorong kontribusi bagi perkembangan pendidikan dan penelitian.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah Subhanahu wa Ta'ala atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi sebagai sebagai salah satu syarat untuk memperoleh gelar Sarjana. Penulis menyadari bahwa proses penyelesaian skripsi ini tidak mungkin berjalan tanpa bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, melalui kesempatan ini, izinkan penulis menyampaikan ucapan terima kasih dan penghargaan yang tulus kepada:

1. Allah Subhanahu wa Ta'ala, atas segala petunjuk, kesehatan, dan kemudahan yang senantiasa dilimpahkan dalam proses penulisan skripsi ini.
2. Kedua Orang Tua tercinta, yang telah melimpahkan kasih sayang, doa tulus yang tidak pernah terputus, pengorbanan, serta dukungan yang menjadi pilar utama penyelesaian studi penulis.
3. Ibu Windha Mega Pradnya D., M.Kom., selaku Dosen Pembimbing yang penuh keikhlasan, dan dedikasi dalam memberikan arahan, bimbingan, dan masukan sejak awal hingga akhir penyusunan skripsi ini.
4. Azzahra Yasmin Syah dan seluruh teman-teman serta sahabat dekat yang selalu memberikan semangat, dukungan, serta masukan positif selama masa perkuliahan dan penulisan skripsi.
5. Seluruh Keluarga Besar yang senantiasa memberikan doa dan perhatian sehingga penulis termotivasi untuk menyelesaikan studi ini dengan baik.

Semoga segala bantuan, dukungan, dan ilmu yang telah diberikan menjadi amal ibadah yang senantiasa mendapatkan balasan terbaik dari Allah Subhanahu wa Ta'ala. Penulis berharap skripsi ini dapat bermanfaat bagi pengembangan ilmu pengetahuan dan pihak-pihak yang berkepentingan.

Yogyakarta, 13 Desember 2025

Raihan Aly Zaky

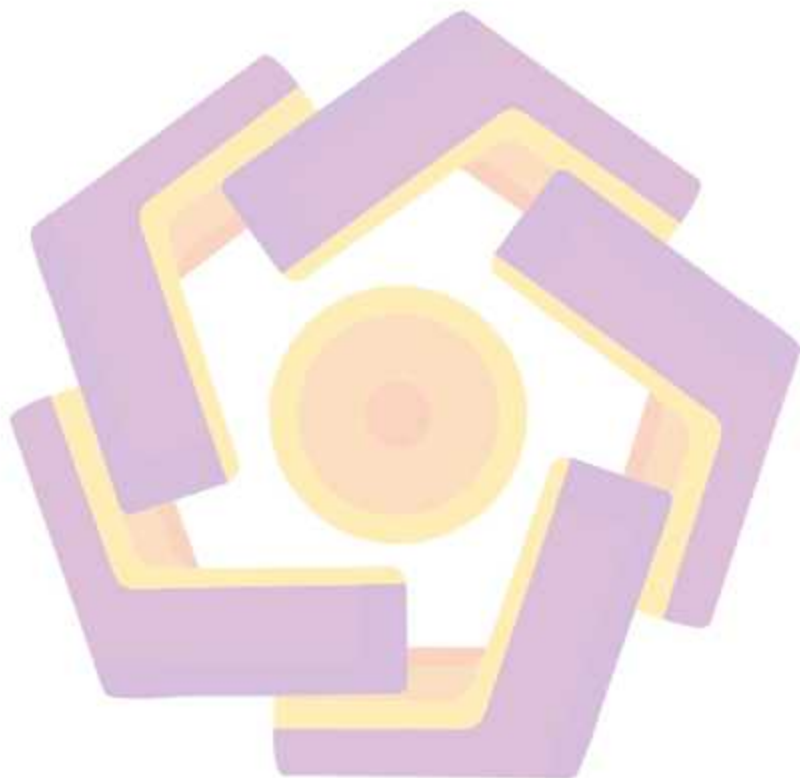
DAFTAR ISI

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.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN.....	xiv
DAFTAR LAMBANG DAN SINGKATAN	xv
DAFTAR ISTILAH	xvi
INTISARI	xviii
ABSTRACT.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	3
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.....	13
2.2.1. Kualitas Air.....	13
2.2.2. Klasifikasi Machine Learning.....	14
2.2.3. Metode Ensemble.....	15
2.2.4. Stacking.....	15
2.2.5. Random Forest.....	16
2.2.6. XGBoost	17
2.2.7. LightGBM.....	19
2.2.8. CatBoost.....	20
2.2.9. Metrik Evaluasi.....	22
2.2.10. RandomizedSearchCV	25
BAB III METODE PENELITIAN	26
3.1 Objek Penelitian.....	26
3.2 Alur Penelitian	26
3.2.1. Muat Dataset.....	27
3.2.2. EDA (Exploratory Data Analysis).....	27
3.2.3. <i>Preprocessing</i>	29
3.2.4. Pelatihan Model Dasar (<i>Base Learner Training</i>).....	31
3.2.5. Stacking.....	31
3.2.6. Evaluasi Model	32
3.3 Alat dan Bahan.....	34
3.3.1. Data Penelitian	34

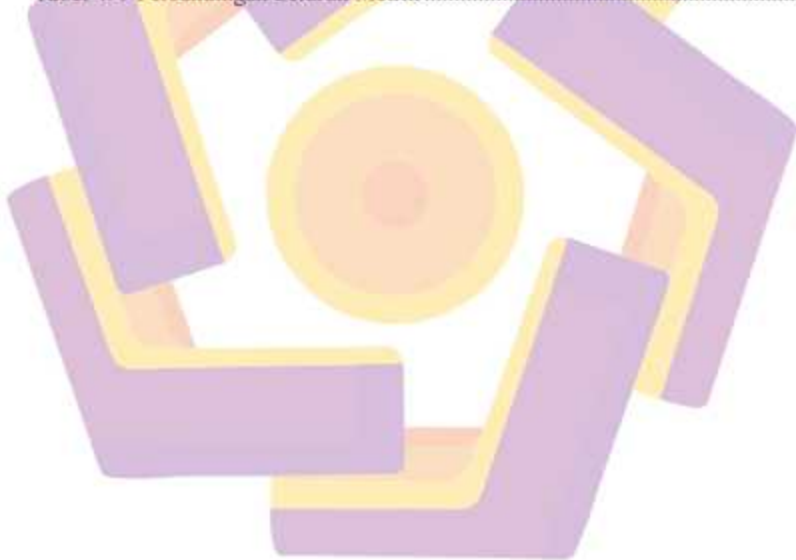
3.3.2. Alat/instrumen.....	35
BAB IV HASIL DAN PEMBAHASAN	36
4.1 Dataset.....	36
4.2 EDA (Exploratory Data Analysis).....	37
4.2.1. Inspeksi Data Awal	38
4.2.2. Cek Statistik Deskriptif.....	39
4.2.3. Pemeriksaan Missing Values	39
4.2.4. Cek Jumlah Variabel Target	40
4.2.5. Pemeriksaan Outliers	41
4.2.6. Analisis Fitur vs Target.....	42
4.2.7. Matriks Korelasi.....	45
4.3 Preprocessing	46
4.3.1. Penanganan Tipe Data (<i>Data Cleaning & Type Handling</i>).....	46
4.3.2. Pembagian Data (<i>Split Data</i>)	48
4.3.3. Penanganan <i>Outliers</i> (IQR Capping).....	49
4.3.4. Scaling (Robust Scaler).....	50
4.3.5. SMOTE	51
4.4 Pelatihan Model Dasar	52
4.5 <i>Stacking</i>	54
4.6 Optimasi <i>Stacking</i>	57
4.7 <i>Hyperparameter Tuning</i>	60
4.8 Perbandingan Evaluasi	64
BAB V PENUTUP	67
5.1 Kesimpulan	67

5.2	Saran	68
REFERENSI		69
LAMPIRAN		74



DAFTAR TABEL

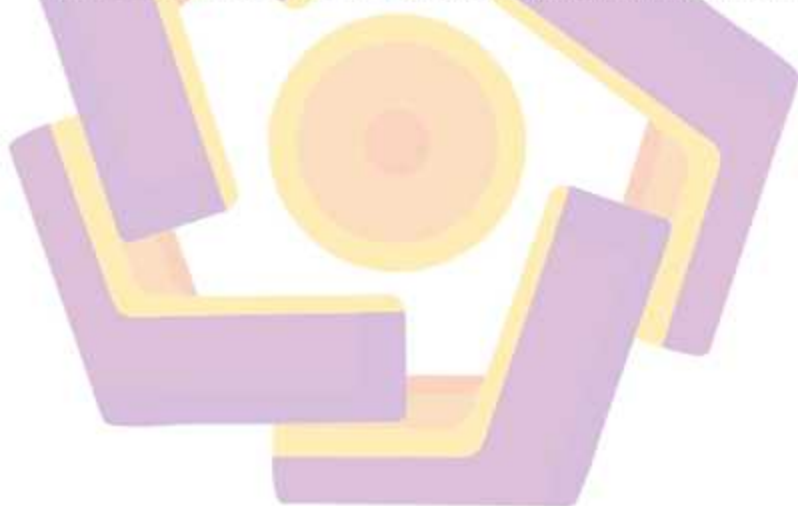
Tabel 4. 1 Dataset <i>Water Quality</i>	36
Tabel 4. 2 Jumlah Target.....	41
Tabel 4. 3 Jumlah Outliers yang Ditemukan.....	42
Tabel 4. 4 Output Data Sebelum dan Sesudah SMOTE	51
Tabel 4. 5 Perbandingan Metrik Base Models.....	53
Tabel 4. 6 Metrik Hasil Stacking.....	56
Tabel 4. 7 Metrik Hasil Optimasi Stacking.....	59
Tabel 4. 8 Metrik Hasil Hyperparameter Tuning.....	62
Tabel 4. 9 Perbandingan Seluruh Metrik	64



DAFTAR GAMBAR

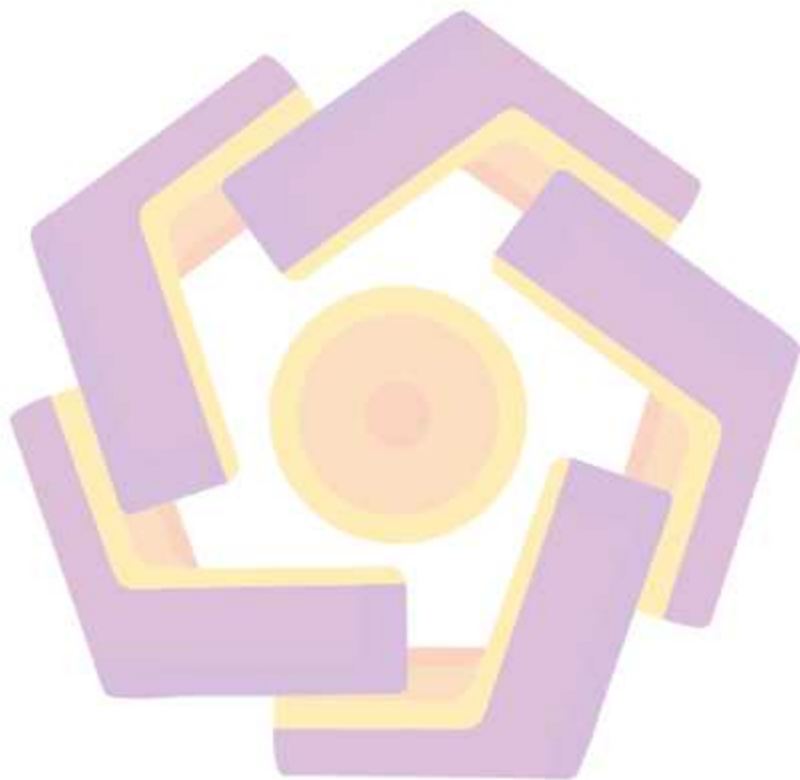
Gambar 2. 1 Ilustrasi Cara Kerja Klasifikasi ML	14
Gambar 2. 2 Arsitektur Metode Ensemble Stacking.....	16
Gambar 2. 3 Arsitektur Algoritma Random Forest	17
Gambar 2. 4 Arsitektur Algoritma XGBoost.....	18
Gambar 2. 5 Arsitektur Algoritma LightGBM (Leaf-wise tree growth)	20
Gambar 2. 6 Arsitektur Algoritma CatBoost	21
Gambar 2. 7 Visualisasi Confusion Matrix.....	22
Gambar 2. 8 Visualisasi Kurva ROC-AUC	25
Gambar 3. 1 Alur Penelitian.....	26
Gambar 3. 2 Dataset Parameter Kualitas Air.....	35
Gambar 4. 1 Hasil Pemeriksaan Data Teratas (Head)	38
Gambar 4. 2 Hasil Pemeriksaan Data Terbawah (Tail)	38
Gambar 4. 3 Tipe Data dari Tiap Fitur.....	38
Gambar 4. 4 Statistik Deskriptif Kolom Numerik	39
Gambar 4. 5 Statistik Deskriptif Kolom Kategori	39
Gambar 4.6 Pemeriksaan Missing Values	40
Gambar 4. 7 Visualisasi Distribusi Target.....	40
Gambar 4. 8 Visualisasi Persebaran Outliers.....	41
Gambar 4. 9 Visualisasi Analisis Fitur vs Target 1	42
Gambar 4. 10 Visualisasi Analisis Fitur vs Target 2	43
Gambar 4. 11 Visualisasi Analisis Fitur vs Target 3	44
Gambar 4. 12 Visualisasi Matriks Korelasi	45
Gambar 4. 13 Pembersihan Data dan Penanganan Tipe Data.....	47
Gambar 4. 14 Pembagian Data Latih dan Data Uji.....	48
Gambar 4. 15 Proses Penanganan Outliers	49
Gambar 4. 16 Proses Scaling dengan RobustScaler	50
Gambar 4. 17 Proses SMOTE pada Data Latih	51
Gambar 4. 18 Pengaturan Parameter Setiap Base Models.....	52
Gambar 4. 19 Confusion Matrix Base Models	53

Gambar 4. 20 Kurva ROC-AUC Base Models	54
Gambar 4. 21 Meta Fitur	55
Gambar 4. 22 Implementasi Classifier dengan RF sebagai Meta-Learner	55
Gambar 4. 23 Confusion Matrix Hasil Prediksi Stacking (RF)	56
Gambar 4. 24 Kurva ROC-AUC Stacking	57
Gambar 4. 25 Implementasi Stacking dengan Parameter Passthrough Aktif	58
Gambar 4. 26 Confusion Matrix Hasil Prediksi Stacking (RF Pass)	59
Gambar 4. 27 Kurva ROC-AUC Optimasi Stacking	60
Gambar 4. 28 Code Implementasi RandomizedSearchCV	61
Gambar 4. 29 Output Implementasi RandomizedSearchCV	61
Gambar 4. 30 Confusion Matrix Hasil Prediksi Setelah Tuning	62
Gambar 4. 31 Kurva ROC-AUC Hyperparameter Tuning	63

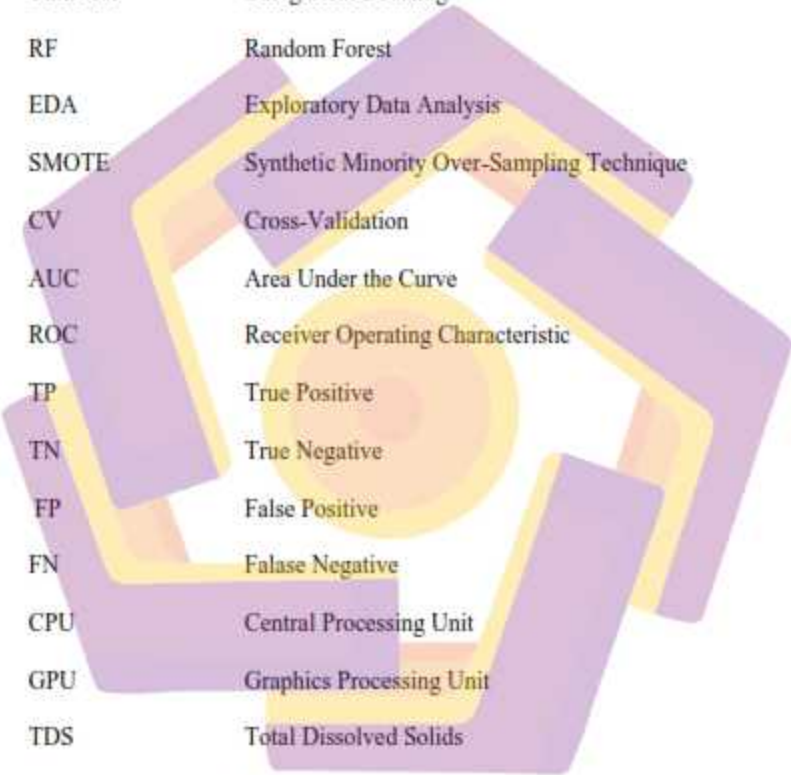


DAFTAR LAMPIRAN

Lampiran 1. Link Dataset.....	35
-------------------------------	----



DAFTAR LAMBANG DAN SINGKATAN

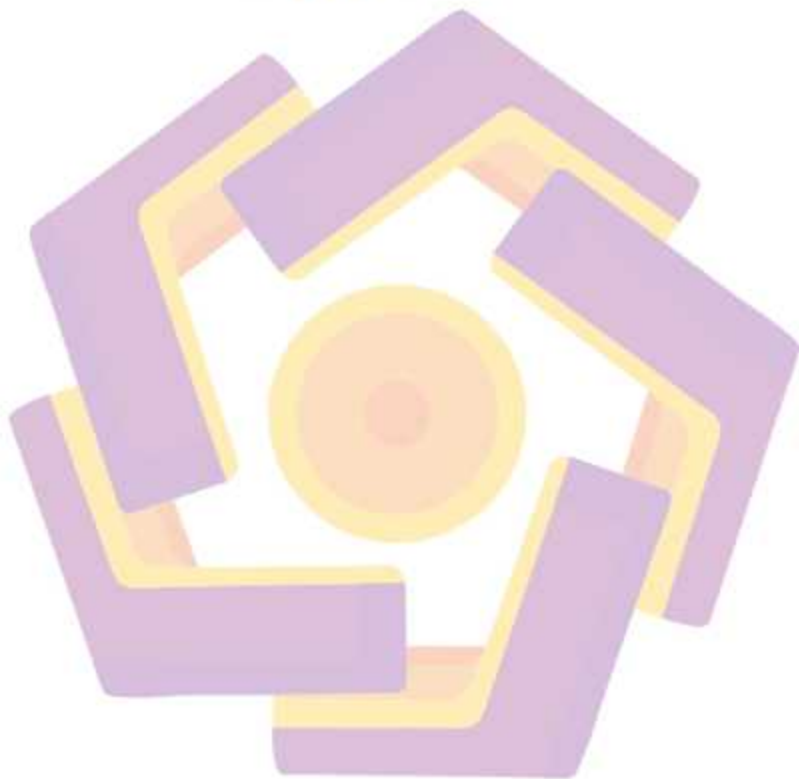


XGBoost	eXtreme Gradient Boost
LightGBM	Light Gradient Boosting Machine
CatBoost	Categorical Boosting
RF	Random Forest
EDA	Exploratory Data Analysis
SMOTE	Synthetic Minority Over-Sampling Technique
CV	Cross-Validation
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
CPU	Central Processing Unit
GPU	Graphics Processing Unit
TDS	Total Dissolved Solids
DO	Dissolved Oxygen
BOD	Biochemical Oxygen Demand
COD	Chemical Oxygen Demand
TSS	Total Suspended Solids

DAFTAR ISTILAH

Stacking	Metode ensemble yang menggabungkan prediksi dari beberapa model dasar (base-learner) menjadi fitur baru (meta-features) untuk dilatih kembali oleh model lain di level lebih tinggi (meta-learner) sehingga menghasilkan prediksi yang lebih akurat.
Meta-Learner	Model tingkat kedua yang mempelajari hubungan antarprediksi model dasar untuk menghasilkan keputusan akhir.
Base-Learner	Model tingkat pertama yang dilatih menggunakan data asli sebelum digabungkan dalam arsitektur stacking.
Meta-Features	Fitur baru berupa output prediksi dari base-learner.
Passthrough	Konfigurasi stacking yang menambahkan fitur asli dataset ke input meta-learner bersama prediksi base-model.
XGBoost	Algoritma boosting yang mengoptimalkan pohon keputusan menggunakan teknik regularisasi dan gradient boosting sehingga lebih cepat dan efisien.
LightGBM	Algoritma boosting yang menggunakan teknik histogram dan leaf-wise growth, sehingga lebih cepat pada dataset besar dan mendukung training berkecepatan tinggi.
CatBoost	Algoritma boosting yang unggul dalam menangani data kategorikal dan mencegah overfitting melalui teknik ordered boosting.
Random Forest	Algoritma ensemble berbasis banyak pohon keputusan yang menggabungkan hasil voting untuk meningkatkan akurasi dan mengurangi overfitting.

Overfitting	Kondisi model terlalu mengikuti pola data latih sehingga performa menurun pada data uji.
Cross-Validation	Teknik evaluasi yang membagi data menjadi beberapa lipatan untuk memastikan model tidak hanya bagus pada satu subset data.



INTISARI

Kualitas air menjadi salah satu isu lingkungan yang berdampak langsung terhadap kesehatan manusia dan ekosistem. Oleh karena itu, diperlukan suatu sistem deteksi kualitas air yang akurat, cepat, dan dapat diandalkan untuk membantu proses pengambilan keputusan. Penelitian ini dilakukan untuk mengatasi permasalahan tersebut. Penelitian ini menggunakan pendekatan metode Ensemble Stacking berbasis algoritma XGBoost, LightGBM, dan CatBoost dengan Random Forest sebagai meta-learner. Selain itu, dilakukan optimasi stacking menggunakan mekanisme passthrough dan hyperparameter tuning berbasis RandomizedSearchCV untuk memperoleh konfigurasi model terbaik.

Hasil penelitian menunjukkan bahwa metode Ensemble Stacking memperoleh hasil akurasi 97,06%, kemudian dengan penambahan parameter passthrough yang telah dioptimasi dengan hyperparameter tuning, memberikan performa terbaik dengan akurasi 97,12% dan precision tertinggi 90,48%. Model ini terbukti mampu menurunkan kesalahan prediksi, terutama False Positive, menjadikannya solusi yang lebih aman untuk penentuan kualitas air. Penelitian ini berkontribusi dalam menyediakan model prediksi yang andal dan dapat dimanfaatkan oleh instansi lingkungan, fasilitas pengolahan air, serta sistem monitoring kualitas air berbasis data.

Kata kunci: Kualitas Air, Klasifikasi, Stacking, Random Forest, Passthrough.

ABSTRACT

Water quality is one of the environmental issues that directly impacts human health and ecosystems. Therefore, an accurate, fast, and reliable water quality detection system is needed to assist in the decision-making process. This study was conducted to address this issue. This study uses an Ensemble Stacking approach based on the XGBoost, LightGBM, and CatBoost algorithms with Random Forest as the meta-learner. In addition, stacking optimization was performed using a passthrough mechanism and RandomizedSearchCV-based hyperparameter tuning to obtain the best model configuration.

The results showed that the Ensemble Stacking method achieved an accuracy of 97.06%. With the addition of the passthrough parameter, which had been optimized with hyperparameter tuning, it delivered the best performance with an accuracy of 97.12% and the highest precision of 90.48%. This model has been proven to reduce prediction errors, especially False Positives, making it a safer solution for determining water quality. This research contributes to providing a reliable prediction model that can be utilized by environmental agencies, water treatment facilities, and data-based water quality monitoring systems.

Keyword: *Water Quality, Classification, Stacking, Random Forest, Passthrough.*