Jurnal Informatika: Jurnal pengembangan IT

Vol. 11, No. 2, 2026
ISSN : 2477-5126, DOI:10.30591/jpit.v11i2.10090

Optimasi Metode Ensemble Stacking Dalam Klasifikasi Kualitas Air Berbasis Algoritma XGBoost, LightGBM, dan CatBoost

Raihan Aly Zaky [1], Windha Mega Pradnya[2]
[1,2]Program Studi Informatika, Univeritas Amikom Yogyakarta, Indonesia
[1]raihanalyzaky@students.amikom.ac.id, [2]windha@amikom.ac.id

Info Artikel

Riwayat Artikel:

Received 2026-01-05
Revised 2026-07-29
Accepted 2026-08-06

Abstract – *Water quality is an important aspect in public health and environmental sustainability, directly influencing decisions on drinking water suitability. Therefore, an accurate classification system is essential for fast and precise decision-making. This study optimizes water quality classification using the Ensemble Stacking method with XGBoost, LightGBM, and CatBoost as base learners, and Random Forest as a meta-learner. The Water Quality Dataset from Kaggle, containing 7,999 samples with 20 features and one target variable (is_safe), was used. Research stages included data exploration, Interquartile Range (IQR) Capping for outlier handling, RobustScaler for normalization, and SMOTE for class balancing. This was followed by building a stacking model with a passthrough mechanism and hyperparameter optimization using RandomizedSearchCV (50 iterations, 5-fold cross-validation). Optimized Random Forest parameters included n_estimators, max_depth, min_samples_split, min_samples_leaf, and max_features to achieve the best configuration with efficient computational cost. Evaluated using Accuracy, Precision, Recall, F1-score, and ROC-AUC metrics, the initial Ensemble Stacking model achieved 97.06% accuracy. However, applying passthrough and hyperparameter optimization improved performance to 97.12% accuracy, 90.48% precision, 83.52% recall, 86.86% F1-score, and 97.66% ROC-AUC. Furthermore, the proposed model effectively reduced False Positives, ensuring safer water quality status determination. Ultimately, this approach shows significant potential for implementation in data-driven water quality monitoring systems, supporting environmental management agencies and water treatment facilities.*

Keywords: *Classification, Water Quality, Pass-Through, Random Forest, Stacking.*

Corresponding Author:

Raihan Aly Zaky

Email:

aihanalyzaky@students.amikom.ac.id

Abstrak – *Kualitas air merupakan aspek penting dalam kesehatan masyarakat dan keberlanjutan lingkungan, yang secara langsung memengaruhi keputusan terkait kelayakan air minum. Oleh karena itu, sistem klasifikasi yang akurat sangat diperlukan untuk mendukung pengambilan keputusan yang cepat dan tepat. Penelitian ini bertujuan mengoptimalkan kinerja klasifikasi kualitas air menggunakan metode Ensemble Stacking dengan algoritma XGBoost, LightGBM, dan CatBoost sebagai base learner, serta Random Forest sebagai meta-learner. Dataset yang digunakan adalah Water Quality dari Kaggle, berisi 7.999 sampel dengan 20 fitur dan satu variabel target (is_safe). Tahapan penelitian meliputi eksplorasi data, penanganan outlier (IQR Capping), normalisasi (RobustScaler), dan penyeimbangan kelas (SMOTE). Tahap selanjutnya adalah pembentukan model stacking dengan mekanisme passthrough serta optimasi hiperparameter menggunakan RandomizedSearchCV (50 iterasi, 5-fold cross-validation). Parameter Random Forest yang dioptimalkan meliputi n_estimators, max_depth, min_samples_split, min_samples_leaf, dan max_features untuk memperoleh konfigurasi terbaik dengan efisiensi komputasi. Dievaluasi menggunakan metrik Accuracy, Precision, Recall, F1-score, dan ROC-AUC, model Ensemble Stacking awal menghasilkan akurasi 97,06%. Penerapan passthrough dan optimasi hiperparameter berhasil meningkatkan performa model menjadi akurasi 97,12%, precision 90,48%, recall 83,52%, F1-score 86,86%, dan ROC-AUC 97,66%. Model yang diusulkan juga terbukti efektif menurunkan jumlah False Positive, sehingga penentuan kelayakan air menjadi lebih aman. Pendekatan ini berpotensi kuat diterapkan pada sistem pemantauan kualitas air berbasis data untuk mendukung instansi pengelola lingkungan dan fasilitas pengolahan air.*

Kata Kunci*: Klasifikasi, Kualitas Air, Passthrough, Random Forest, Stacking*

I. PENDAHULUAN

Kualitas air merupakan faktor fundamental dalam menjaga kesejahteraan manusia dan kelangsungan ekosistem, mengingat ketersediaan air bersih yang aman berperan penting dalam mendukung aktivitas sosial dan ekonomi. Kualitas air ditentukan oleh berbagai parameter fisika, kimia, dan biologi yang mencerminkan kondisi suatu sumber air serta kelayakannya untuk digunakan sesuai peruntukan tertentu [1]. Kompleksitas dan keterkaitan antar parameter tersebut menyebabkan proses evaluasi kualitas air menjadi tidak sederhana, terutama ketika melibatkan volume data yang besar dan beragam. Di Indonesia, hasil pemantauan mutu air menunjukkan adanya upaya perbaikan, seperti peningkatan mutu pada sungai, namun permasalahan fragmentasi data dan ketimpangan pengelolaan kualitas air antarwilayah masih menjadi tantangan yang signifikan [2][3]. Selain itu, perbedaan karakteristik geografis dan

aktivitas antropogenik turut mempengaruhi variasi kualitas air, sehingga diperlukan pendekatan analitis yang adaptif dan berbasis data [4]. Kondisi ini menuntut adanya sistem klasifikasi kualitas air yang lebih akurat, efisien, dan mampu memproses data dalam skala besar.

Pengukuran kualitas air secara konvensional umumnya dilakukan melalui pengujian laboratorium yang kompleks, memerlukan biaya tinggi, serta membutuhkan waktu yang relatif lama. Kondisi tersebut membatasi frekuensi pemantauan dan ketersediaan informasi kualitas air secara real time, khususnya pada wilayah dengan keterbatasan sumber daya. Seiring dengan perkembangan teknologi, pendekatan machine learning mulai banyak diterapkan untuk membantu proses evaluasi dan klasifikasi kualitas air secara lebih cepat dan efisien [5]. Pendekatan ini memungkinkan pemanfaatan pola tersembunyi dalam data multidimensi yang sulit diidentifikasi melalui metode konvensional**.** Namun, sejumlah penelitian menunjukkan bahwa penggunaan model machine learning tunggal masih memiliki keterbatasan performa. Studi yang menerapkan Random Forest untuk klasifikasi kelayakan air minum hanya menghasilkan akurasi sebesar 66,6% [6]. Penelitian lain yang membandingkan beberapa algoritma boosting melaporkan akurasi 68% untuk CatBoost, 60% untuk Gradient Boosting, dan 58% untuk AdaBoost [7]. Bahkan model XGBoost yang dikenal memiliki performa tinggi masih menunjukkan hasil yang belum optimal dengan akurasi sebesar 82,29% [8]. Temuan-temuan tersebut mengindikasikan perlunya strategi pemodelan yang lebih robust guna meningkatkan akurasi dan stabilitas klasifikasi kualitas air.

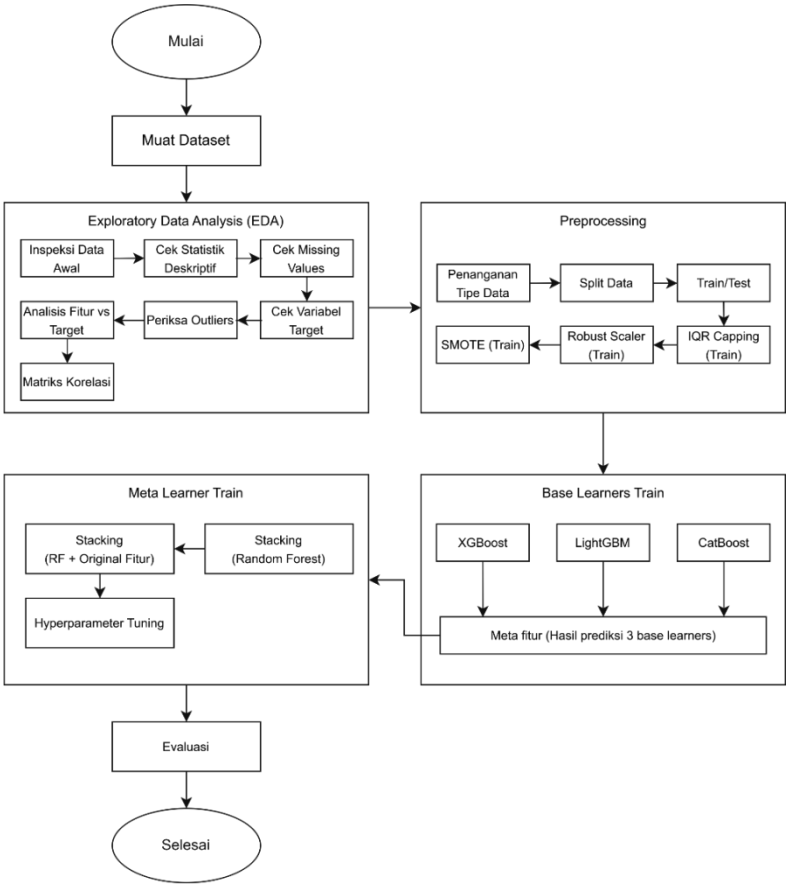
Keterbatasan tersebut mendorong pengembangan pendekatan ensemble learning, khususnya metode stacking, yang mengkombinasikan beberapa model untuk meningkatkan kinerja klasifikasi. Pendekatan ini dirancang untuk meminimalkan bias dan varians yang kerap muncul pada penggunaan algoritma tunggal. Penelitian terdahulu menunjukkan bahwa metode ensemble mampu memberikan performa yang lebih baik dibandingkan model tunggal dalam klasifikasi mutu air [9]. Metode ensemble stacking secara khusus terbukti efektif dalam meningkatkan metrik akurasi, presisi, dan recall dengan menggabungkan prediksi dari beberapa base-learner yang kemudian dipelajari oleh sebuah meta-learner [10][11]. Dengan mekanisme pembelajaran bertingkat, model *stacking* dapat menangkap pola kompleks yang tidak dapat diidentifikasi secara optimal oleh satu algoritma saja. Pendekatan ini memungkinkan pemanfaatan keunggulan masing-masing algoritma sehingga menghasilkan model yang lebih robust [12] [13].

Algoritma gradient boosting modern seperti XGBoost, LightGBM, dan CatBoost dikenal memiliki kinerja yang unggul dalam menangani data kompleks dan berskala besar. Ketiga algoritma tersebut memiliki karakteristik pembelajaran yang saling melengkapi dalam menangani variasi distribusi data dan interaksi nonlinier antar fitur, yang terbukti efektif ketika dikombinasikan dalam model *stacking ensemble* untuk meningkatkan akurasi klasifikasi dibanding model tunggal [14]. Pendekatan serupa juga menunjukkan bahwa integrasi beberapa algoritma booster dalam *stacking* mampu menghasilkan prediksi yang lebih stabil dan akurat, karena memanfaatkan kekuatan unik masing-masing algoritma dalam menangkap berbagai pola data [15]. XGBoost menawarkan efisiensi komputasi dan mekanisme regularisasi yang efektif dalam mengendalikan overfitting [16]. LightGBM mengandalkan strategi leaf-wise growth dan Gradient-based One-Side Sampling (GOSS) untuk meningkatkan kecepatan pelatihan dan efisiensi memori [17]. Sementara itu, CatBoost unggul dalam menangani fitur kategorikal keunggulan dalam menangani fitur kategorikal secara langsung tanpa memerlukan pra pemrosesan tambahan [18]. Kombinasi ketiga algoritma ini sebagai *base-learner* diharapkan mampu memperkaya representasi pola pada data kualitas air. Untuk mengagregasi prediksi dari ketiga algoritma tersebut, Random Forest dipilih sebagai meta-learner karena kemampuannya dalam menjaga stabilitas model dan meminimalkan overfitting melalui mekanisme ensemble berbasis voting [19].

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menerapkan dan mengoptimasi metode *ensemble stacking* dalam klasifikasi kualitas air dengan menggunakan XGBoost, LightGBM, dan CatBoost sebagai *base-learner* serta Random Forest sebagai *meta-learner*. Pendekatan ini dipilih karena mampu mengkombinasikan keunggulan masing-masing algoritma dalam menangani pola non linier dan kompleks pada data kualitas air melalui mekanisme pembelajaran bertingkat. Dibandingkan Logistic Regression yang bersifat linier, Random Forest lebih mampu memodelkan interaksi kompleks antar fitur hasil prediksi serta memiliki ketahanan yang lebih baik terhadap overfitting melalui mekanisme bootstrap aggregation dan pemilihan fitur secara acak pada setiap pohon keputusan. Karakteristik tersebut menjadikan Random Forest lebih sesuai sebagai penggabung (aggregator) pada penelitian ini, terutama karena seluruh base learner menghasilkan probabilitas prediksi dengan karakteristik yang berbeda sehingga diperlukan meta-learner yang mampu memanfaatkan keragaman tersebut secara optimal. Optimasi dilakukan melalui penerapan mekanisme *passthrough* atau penambahan fitur asli dan *hyperparameter tuning* berbasis *RandomizedSearchCV* bekerja dengan cara memilih subset acak dari kombinasi parameter yang tersedia [20], guna memperoleh konfigurasi model terbaik sekaligus meningkatkan kemampuan generalisasi. Kebaruan penelitian ini terletak pada integrasi tiga algoritma *gradient boosting* unggulan dalam kerangka *stacking* yang dioptimasi secara sistematis untuk klasifikasi kualitas air, dengan fokus pada peningkatan akurasi dan penurunan kesalahan prediksi, khususnya *False Positive*, sehingga menghasilkan model yang lebih andal untuk mendukung sistem pemantauan kualitas air berbasis data.

II. METODE

Penelitian ini dilaksanakan melalui serangkaian tahapan yang tersusun secara sistematis untuk memastikan proses analisis dan pemodelan berjalan terstruktur dan terukur. Alur penelitian diawali dengan pengumpulan dan pemahaman data kualitas air, di mana dataset yang digunakan diperoleh dari platform Kaggle sebagai sumber data sekunder, dilanjutkan dengan tahap pra pemrosesan data untuk memastikan kualitas dan kesiapan data sebelum digunakan dalam pemodelan. Selanjutnya, dilakukan pembangunan model klasifikasi menggunakan pendekatan *ensemble stacking* yang melibatkan beberapa algoritma *base-learner* dan *meta-learner*, disertai dengan proses optimasi parameter guna memperoleh kinerja model yang optimal. Tahap akhir penelitian mencakup evaluasi performa model berdasarkan metrik klasifikasi yang relevan serta analisis hasil untuk menilai efektivitas pendekatan yang diusulkan. Keseluruhan tahapan tersebut dirangkum dalam suatu diagram alur proses penelitian sebagaimana ditampilkan pada Gambar 1.



Gambar 1. Alur Penelitian

A. Pengumpulan Data

Tahap awal penelitian dilakukan dengan mengumpulkan dataset kualitas air yang bersumber dari platform Kaggle sebagai data sekunder. Dataset ini berisi sejumlah parameter kualitas air yang digunakan sebagai dasar dalam proses klasifikasi. Dataset ini mencakup total 7.999 sampel data yang terstruktur dalam 21 atribut. Komposisi tersebut meliputi 20 fitur bertipe numerik yang merepresentasikan parameter kualitas air, serta satu label target bernama is_safe. Variabel target ini berfungsi sebagai kelas acuan dalam menentukan status kualitas air, apakah aman atau tidak aman. Gambar 2 menampilkan dataset yang digunakan dalam penelitian ini

	aluminium	ammonia	arsenic	barium	cadmium	chloramine	chromium	copper	flouride	bacteria	viruses	lead	nitrates	nitrites	mercury	perchlorate	radium	selenium	silver	uranium	is_safe
2	1.65	9.08	0.04	2.85	0.007	0.35	0.83	0.17	0.05	0.2	0	0.054	16.08	1.13	0.007	37.75	6.78	0.08	0.34	0.02	1
3	2.32	21.16	0.01	3.31	0.002	5.28	0.68	0.66	0.9	0.65	0.65	0.1	2.01	1.93	0.003	32.26	3.21	0.08	0.27	0.05	1
4	1.01	14.02	0.04	0.58	0.008	4.24	0.53	0.02	0.99	0.05	0.003	0.078	14.16	1.11	0.006	50.28	7.07	0.07	0.44	0.01	0
5	1.36	11.33	0.04	2.96	0.001	7.23	0.03	1.66	1.08	0.71	0.71	0.016	1.41	1.29	0.004	9.12	1.72	0.02	0.45	0.05	1
6	0.92	24.33	0.03	0.2	0.006	2.67	0.69	0.57	0.61	0.13	0.001	0.117	6.74	1.11	0.003	16.9	2.41	0.02	0.06	0.02	1
7	0.94	14.47	0.03	2.88	0.003	0.8	0.43	1.38	0.11	0.67	0.67	0.135	9.75	1.89	0.006	27.17	5.42	0.08	0.19	0.02	1
8	2.36	5.6	0.01	1.35	0.004	1.28	0.62	1.88	0.33	0.13	0.007	0.021	18.6	1.78	0.007	45.34	2.84	0.1	0.24	0.08	0
9	3.93	19.87	0.04	0.66	0.001	6.22	0.1	1.86	0.86	0.16	0.005	0.197	13.65	1.81	0.001	53.35	7.24	0.08	0.08	0.07	0
10	0.6	24.58	0.01	0.71	0.005	3.14	0.77	1.45	0.98	0.35	0.002	0.167	14.66	1.84	0.004	23.43	4.99	0.08	0.25	0.08	1
11	0.22	16.76	0.02	1.37	0.007	6.4	0.49	0.82	1.24	0.83	0.83	0.109	4.79	1.46	0.01	30.42	0.08	0.03	0.31	0.01	1
12	3.27	3.6	0.001	2.69	0.005	5.75	0.15	0.6	1.29	0.04	0.008	0.145	8.47	1.25	0.006	55.4	7.8	0.05	0.33	0.06	0
13	1.35	21.96	0.04	0.84	0.002	0.1	0.76	0.17	0.58	0.52	0.52	0.011	18.4	1.49	0.009	21.52	1.3	0.08	0.48	0.08	1
14	1.88	19.26	0.02	2.78	0.008	0.05	0.42	1	0.09	0.91	0.91	0.103	4.37	1.95	0.006	22.12	1.97	0.03	0.06	0.05	1
15	4.93	23.98	0.04	3.05	0.008	0.7	0.51	1.35	1.07	0.7	0.7	0.101	1.16	1.11	0.008	26.8	5.58	0.09	0.38	0.03	1
16	2.89	18.82	0.05	3.77	0.008	5.99	0.54	0.79	0.54	0.2	0.009	0.126	17.56	1.82	0.009	17.54	4.33	0.1	0.05	0.02	1
17	0.61	2.41	0.03	0.59	0.002	1.94	0.77	1.54	0.62	0.23	0.001	0.017	1.99	1.08	0.007	11.16	0.98	0.01	0.47	0.03	1
18	3.47	15.84	0.02	0.06	0.001	5.29	0.47	1.08	1.43	0.89	0.89	0.08	1.91	1.2	0.008	0.18	6.89	0.06	0.12	0.08	1
19	2.11	17.03	0.02	0.88	0.009	7.78	0.88	1.15	0.34	0.85	0.85	0.065	17.86	1.53	0.003	19.4	1.14	0.1	0.4	0.01	1
20	4.88	26.94	0.02	0.36	0.001	1.21	0.68	0.71	0.99	0.75	0.75	0.071	0.31	1.22	0.002	56.7	1	0	0.41	0.05	0

Gambar 2. Dataset Kualitas Air

Dataset ini memiliki beberapa keterbatasan antara lain distribusi kelas yang tidak seimbang menyebabkan model cenderung memprediksi kelas mayoritas apabila tidak dilakukan teknik penyeimbangan data. Oleh karena itu, penelitian ini menerapkan SMOTE untuk meningkatkan representasi kelas minoritas selama proses pelatihan model.

B. Eksplorasi Data

Pada tahap ini dilakukan analisis awal terhadap struktur dan karakteristik data untuk memahami distribusi fitur, tipe data, serta potensi permasalahan seperti nilai hilang dan ketidakseimbangan kelas. Proses yang dilakukan meliputi.

1. Inspeksi Data Awal

Langkah ini dilakukan dengan meninjau sebagian data awal dan akhir serta tipe data setiap fitur untuk memastikan data berhasil dimuat dengan benar. Proses ini bertujuan memverifikasi kesesuaian tipe data dan format nilai.

2. Cek Statistik Deskriptif

Analisis statistik deskriptif diterapkan untuk merangkum karakteristik data menggunakan ukuran utama seperti mean, median, standar deviasi, nilai minimum, dan maksimum. Analisis ini memberikan gambaran tendensi sentral dan variasi data pada parameter fisik dan kimia air, sehingga rentang nilai setiap fitur dapat dipahami dengan lebih jelas.

3. Pemeriksaan Missing Values

Identifikasi keberadaan nilai yang hilang (*null*) pada setiap atribut dilakukan secara menyeluruh. Deteksi ini krusial untuk menentukan langkah penanganan selanjutnya.

4. Cek Jumlah Variabel Target

Analisis distribusi frekuensi pada kolom target, dilakukan untuk mengetahui proporsi kelas data. Langkah ini berfungsi mendeteksi adanya ketimpangan kelas.

5. Pemeriksaan Outliers

Pemeriksaan pencilan (*outliers*) dilakukan untuk menemukan nilai ekstrem yang menyimpang jauh dari sebaran data normal.

6. Analisis Fitur vs Target

Evaluasi hubungan antara masing-masing fitur independen terhadap variabel target dilakukan untuk melihat pola pembeda. Analisis ini membantu memahami parameter mana yang memiliki pengaruh signifikan terhadap klasifikasi kualitas air.

7. Matriks Korelasi

Pembuatan matriks korelasi dilakukan untuk mengukur kekuatan hubungan linier antar variabel. Analisis ini berguna untuk mendeteksi multikolinearitas, yaitu korelasi tinggi antar fitur yang dapat mempengaruhi efisiensi komputasi model.

C. Preprocessing

Preprocessing data merupakan tahapan krusial untuk mengubah data mentah menjadi format yang bersih dan terstruktur agar dapat diproses secara optimal oleh algoritma *machine learning*. Proses ini bertujuan untuk meningkatkan kualitas data. Alur pra-pemrosesan dirancang secara bertahap yang dilakukan penuh pada data latih.

1. Penanganan Tipe Data

Langkah ini difokuskan untuk memastikan integritas tipe data pada seluruh variabel dengan menyesuaikan setiap fitur sesuai karakteristik datanya. Proses ini bertujuan mencegah kesalahan komputasi serta menjamin data dapat diproses secara optimal pada tahap pemodelan selanjutnya.

2. Split Data

Dilakukan pembagian dataset menjadi dua bagian terpisah, yaitu data latih (*training set*) dan data uji (*testing set*). Pemisahan ini bertujuan untuk mensimulasikan evaluasi model pada data yang belum pernah dilihat sebelumnya, guna mengukur kemampuan generalisasi model secara objektif.

3. Train/Test Set

Hasil dari proses *split data* menghasilkan dua subset data. Data latih akan digunakan sebagai landasan pelatihan bagi model untuk mengenali pola, sedangkan data uji disimpan secara terisolasi dan hanya digunakan pada tahap akhir evaluasi. Penelitian stacking ensemble modern menegaskan pentingnya pembagian data ini dan penggunaan data uji yang terpisah untuk memastikan estimasi performa model yang tidak bias [21]. Penting untuk dicatat bahwa langkah pra-pemrosesan selanjutnya hanya diterapkan pada data latih untuk mencegah kebocoran data

4. Penanganan Outliers

Penanganan pencilan diterapkan secara eksklusif pada data latih untuk mencegah bias evaluasi. Nilai ekstrem yang berpotensi mempengaruhi pembentukan batas keputusan model dibatasi menggunakan metode *Interquartile Range (IQR) capping*.

5. Scaling

Penyekalaan fitur diterapkan pada data latih menggunakan *Robust Scaler* untuk menyamakan rentang nilai antar variabel dan mengurangi pengaruh pencilan. RobustScaler digunakan atas dasar karakteristik dataset kualitas air yang mengandung sejumlah nilai ekstrem (outlier) pada beberapa parameter, seperti aluminium, arsenic, dan nitrites. Berbeda dengan Min-Max Scaling yang menggunakan nilai minimum dan maksimum, maupun StandardScaler yang bergantung pada rata-rata dan simpangan baku, RobustScaler melakukan transformasi berdasarkan median dan Interquartile Range (IQR) sehingga lebih tahan terhadap pengaruh pencilan. Karakteristik tersebut menjadikan RobustScaler lebih sesuai untuk mempertahankan distribusi utama data tanpa memberikan pengaruh berlebihan dari nilai-nilai ekstrem. Meskipun algoritma gradient boosting relatif invarian terhadap skala, normalisasi ini membantu meningkatkan stabilitas numerik dan efisiensi proses pelatihan [22]. Selain itu, penggunaan RobustScaler juga membantu menghasilkan proses pembelajaran yang lebih konsisten pada algoritma ensemble yang digunakan dalam penelitian ini. Parameter skala yang diperoleh dari data latih selanjutnya diterapkan secara konsisten pada data uji.

6. SMOTE

Pada tahap akhir pra-pemrosesan data latih, teknik *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan untuk mengatasi ketidakseimbangan kelas dengan membangkitkan sampel sintetis pada kelas minoritas [23], sehingga model dapat mempelajari pola kedua kelas secara lebih seimbang dan mengurangi bias terhadap kelas mayoritas [24].

D. Pelatihan Model Dasar

Tahap ini berfokus pada pelatihan model dasar (Level-0) menggunakan tiga algoritma *Gradient Boosting*: XGBoost, LightGBM, dan CatBoost. Penalaan *hyperparameter* dilakukan secara manual pada setiap model untuk mengoptimalkan generalisasi dan memitigasi *overfitting*. Evaluasi kinerja individual dilakukan terlebih dahulu untuk memvalidasi efektivitas tiap algoritma sebelum diintegrasikan ke dalam skema *stacking*.

E. Stacking

Pada tahap ini, luaran dari model Level-0 tidak diperlakukan sebagai hasil akhir, melainkan dikonsolidasikan sebagai input bagi algoritma Random Forest di Level-1. Sebagai *meta-learner*, Random Forest bertugas mengagregasi keputusan melalui pendekatan bagging yang bekerja secara paralel [25], serta mempelajari korelasi antara prediksi dasar dengan label aktual untuk mengoptimalkan generalisasi model secara keseluruhan. Untuk memperkaya konteks informasi bagi meta-learner, penelitian ini mengimplementasikan mekanisme passthrough, di mana data fitur original diproses bersamaan dengan hasil prediksi model dasar. Selain itu, optimasi kinerja dilakukan melalui hyperparameter

tuning pada struktur stacking untuk menjamin stabilitas dan presisi model akhir. Optimasi hyperparameter dilakukan menggunakan RandomizedSearchCV sebanyak 50 kombinasi parameter (n_iter = 50) dengan 5-fold cross validation. Pendekatan ini dipilih karena memiliki efisiensi komputasi yang lebih baik dibandingkan Grid Search namun tetap mampu menemukan kombinasi parameter yang mendekati optimum pada ruang pencarian yang luas.

F. Evaluasi

Setelah tahapan pelatihan dan pembangunan arsitektur *Stacking Ensemble*, penelitian memasuki fase pengukuran hasil untuk mengevaluasi kemampuan generalisasi model pada data baru. Evaluasi dilakukan secara kuantitatif menggunakan data uji guna memastikan performa model tetap konsisten dan objektif pada kondisi riil, tidak hanya pada data pelatihan.Evaluasi kinerja model merupakan tahap akhir yang krusial untuk memvalidasi keandalan metode *Stacking Ensemble* dalam mengklasifikasikan kualitas air. Analisis ini didasarkan pada

1. *Akurasi (Accuracy)*

   Merupakan metrik paling mendasar yang merepresentasikan rasio total prediksi yang benar (baik positif maupun negatif) terhadap keseluruhan jumlah data yang diuji.

2. *Presisi (Precision)*

   Mengukur tingkat ketepatan atau reliabilitas model saat memprediksi kelas positif. Metrik ini menunjukkan seberapa besar proporsi data yang diklasifikasikan sistem sebagai "aman" benar-benar merupakan air yang aman dalam kenyataannya.

3. *Sensitivitas (Recall)*

   Mengukur kemampuan model dalam menemukan kembali atau mencakup seluruh data yang sebenarnya positif. Metrik ini berfokus pada seberapa banyak sampel air aman yang berhasil dideteksi oleh sistem dari total sampel aman yang ada di dataset.

4. *F1-Score*

   Merupakan rata-rata harmonik yang menggabungkan nilai Presisi dan Recall. Metrik ini penting pada *imbalanced dataset* karena memberikan evaluasi yang lebih seimbang antara ketepatan prediksi dan kelengkapan deteksi.

5. *ROC-AUC*

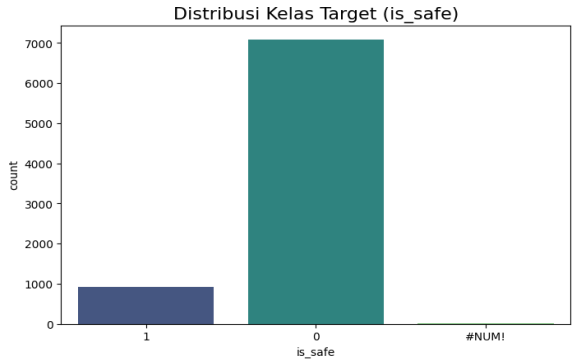
   Mengukur kemampuan model dalam memisahkan kelas positif dan kelas negatif pada beragam nilai ambang keputusan. AUC merefleksikan peluang bahwa model memberikan skor lebih tinggi pada sampel positif dibandingkan sampel negatif, sehingga nilai AUC yang semakin besar menunjukkan kinerja model yang semakin optimal dalam membedakan kedua kelas.

III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian yang diawali dengan *Exploratory Data Analysis* (EDA), tahap preprocessing data, serta penerapan metode *stacking* pada dataset kualitas air.

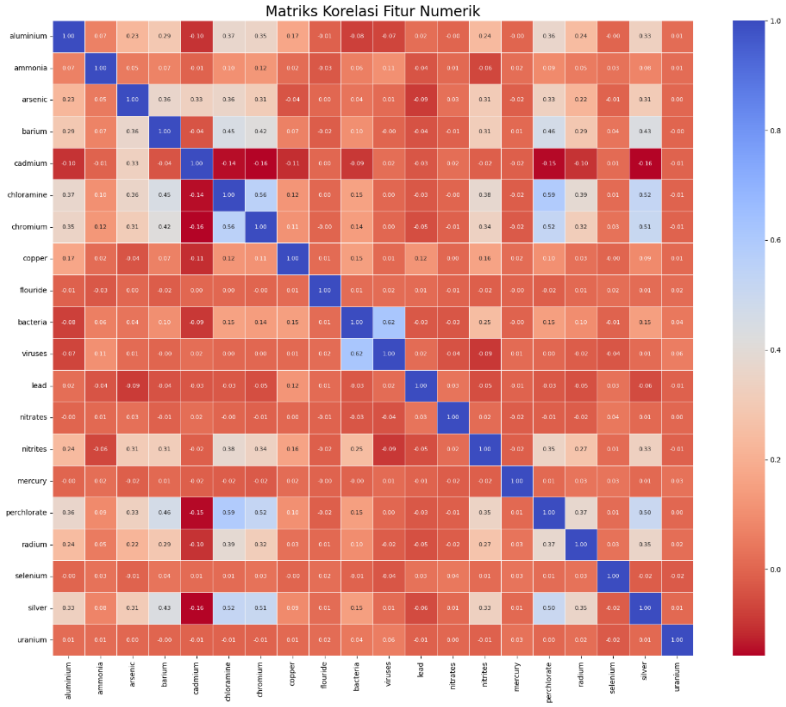
A. Hasil Exploratory Data Analysis (EDA)

Tahap Exploratory Data Analysis (EDA) dilakukan untuk memperoleh pemahaman awal terhadap karakteristik dataset kualitas air sebelum memasuki tahap pemodelan. Dataset terdiri dari 21 fitur yang merepresentasikan parameter fisika dan kimia air, dengan satu variabel target biner yang mengklasifikasikan kualitas air sebagai safe dan not safe. Inspeksi awal data menunjukkan bahwa struktur dataset konsisten, dengan tipe data numerik dan kategorikal yang terdefinisi dengan baik, di mana fitur *ammonia* direpresentasikan sebagai variabel kategorikal berdasarkan rentang atau tingkat konsentrasi tertentu. Analisis statistik deskriptif mengindikasikan adanya variasi rentang nilai yang cukup besar pada beberapa parameter kualitas air, yang mencerminkan heterogenitas kondisi sampel air. Pemeriksaan missing values menunjukkan bahwa dataset relatif bersih dan tidak mengandung nilai hilang yang signifikan, sehingga tidak diperlukan imputasi tambahan. Distribusi kelas target memperlihatkan ketidakseimbangan kelas, di mana salah satu kelas memiliki jumlah observasi yang lebih dominan. Kondisi ini berpotensi menimbulkan bias pada proses pelatihan model dan menjadi dasar penerapan teknik penyeimbangan data pada tahap preprocessing, distribusi target divisualisasikan pada Gambar 3.



Gambar 3. Distribusi Target

Selanjutnya, analisis outlier mengidentifikasi adanya nilai ekstrem pada beberapa fitur numerik, khususnya *alumunium, arsenic,* dan *nitrites*, yang berpotensi mempengaruhi stabilitas model jika tidak ditangani. Analisis hubungan fitur terhadap variabel target menunjukkan perbedaan pola distribusi pada beberapa parameter kualitas air di setiap kelas. Selain itu, analisis korelasi antar fitur menunjukkan tidak adanya korelasi linear yang sangat tinggi, sehingga risiko multikolinearitas relatif rendah dan mendukung penggunaan algoritma berbasis pohon keputusan, sebagaimana ditunjukkan pada Gambar 4. Secara keseluruhan, hasil EDA menegaskan bahwa dataset cukup kompleks untuk diklasifikasikan menggunakan pendekatan ensemble dan memerlukan preprocessing lanjutan sebelum pemodelan.

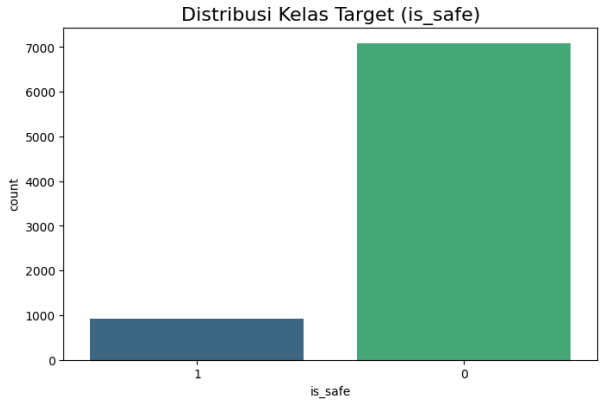


Gambar 4. Matriks Korelasi

B. Hasil Preprocessing

Tahap preprocessing dilakukan untuk meningkatkan kualitas dataset sebelum proses pelatihan model machine learning. Proses ini bertujuan memastikan data berada dalam kondisi bersih, konsisten, seimbang, dan siap digunakan oleh algoritma ensemble stacking.

Langkah awal preprocessing difokuskan pada penanganan tipe data dan pembersihan nilai anomali pada dataset, khususnya pada variabel target *is_safe*. Setelah dilakukan konversi nilai non-numerik ke dalam format numerik serta penghapusan baris data yang tidak valid, distribusi kelas target menjadi lebih konsisten sebagaimana ditunjukkan pada Gambar 5. Selain itu, fitur *ammonia* yang sebelumnya bertipe objek berhasil dikonversi menjadi numerik, dengan nilai hilang ditangani menggunakan imputasi median untuk menjaga kelengkapan dan konsistensi data sebelum tahap pemrosesan selanjutnya.



Gambar 5. Hasil Setelah Penanganan Tipe Data dan Nilai Anomali

Penanganan outlier diterapkan pada data latih menggunakan metode *Interquartile Range* (IQR) Capping untuk membatasi pengaruh nilai ekstrem yang berpotensi mempengaruhi pembentukan *decision boundary* model. Batas bawah dan batas atas ditentukan berdasarkan distribusi kuartil pada data latih, kemudian digunakan sebagai nilai ambang untuk melakukan pembatasan pada fitur yang teridentifikasi mengandung outlier. Hasil penentuan batas capping untuk masing-masing fitur disajikan pada Tabel 1.

TABEL 1
HASIL PENANGANAN OUTLIER

Fitur	Lower Bound	Upper Bound
alumunium	–0.3350	0.6650
arsenic	–0.0750	0.2050
nitrites	–0.1400	2.9000

Proses feature scaling diterapkan menggunakan metode *RobustScaler* guna menyesuaikan rentang nilai antarfitur serta meningkatkan stabilitas numerik selama pelatihan. Meskipun algoritma boosting relatif tahan terhadap skala data, normalisasi tetap memberikan keuntungan dalam mempercepat konvergensi dan meningkatkan konsistensi performa model.

Sebagai tahap akhir preprocessing, diterapkan *Synthetic Minority Over-sampling Technique* (SMOTE) pada data latih untuk menangani ketidakseimbangan kelas. Perbandingan distribusi kelas sebelum dan sesudah penerapan SMOTE, sebagaimana ditunjukkan pada Tabel 2, memperlihatkan bahwa jumlah sampel pada kelas minoritas berhasil ditingkatkan hingga mendekati seimbang dengan kelas mayoritas.

TABEL 2
HASIL SMOTE

Sebelum SMOTE	Setelah SMOTE
0 = 5666	0 = 5666
1 = 730	1 = 5666

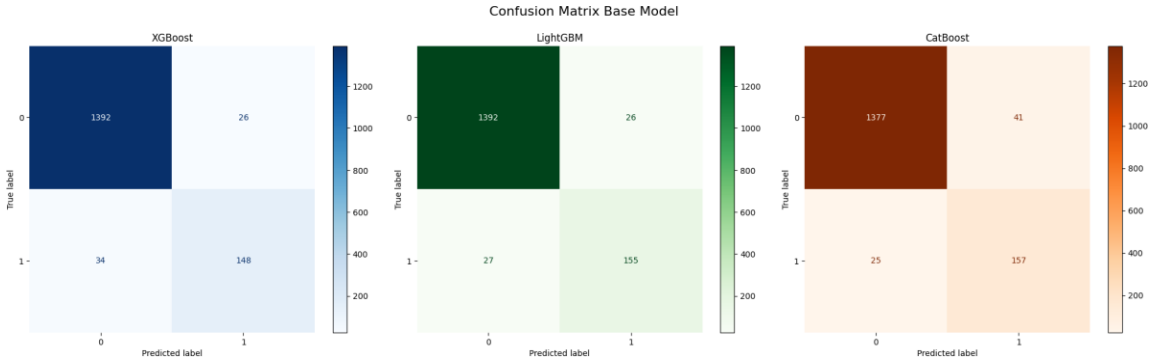
C. Hasil Pelatihan Model Dasar

Pada tahap pelatihan model dasar (Level-0), digunakan tiga algoritma berbasis gradient boosting, yaitu XGBoost, LightGBM, dan CatBoost. Masing-masing model dilatih secara independen menggunakan data latih yang telah melalui tahap preprocessing dan penyeimbangan kelas dengan SMOTE, dengan penyetelan parameter manual untuk memperoleh kinerja yang optimal. Hasil evaluasi performa model dasar yang disajikan pada Tabel 3 menunjukkan bahwa ketiga algoritma mampu mencapai akurasi tinggi, dengan LightGBM memperoleh performa paling konsisten berdasarkan nilai accuracy, precision, recall, dan F1-score.

TABEL 3
HASIL EVALUASI BASE LEARNER

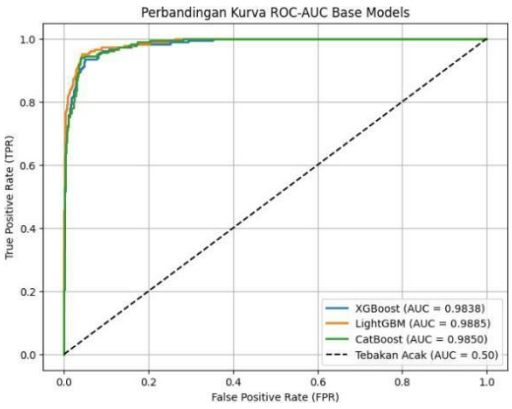
Model	Accuracy	Precision	Recall	F1-Score
XGBoost	0.9625	0.8506	0.8132	0.8315
LightGBM	0.9669	0.8564	0.8516	0.8540
CatBoost	0.9587	0.7929	0.8626	0.8263

Analisis *confusion matrix* pada ketiga model dasar menunjukkan perbedaan karakteristik prediksi antar kelas. Berdasarkan Gambar 6, XGBoost dan LightGBM mampu mengklasifikasikan kelas negatif dengan tingkat ketepatan yang tinggi, ditunjukkan oleh dominasi nilai *true negative* dan kesalahan prediksi yang relatif rendah. LightGBM menghasilkan jumlah *true positive* tertinggi, yang mencerminkan kemampuannya dalam mengenali kelas minoritas. Sementara itu, CatBoost menunjukkan peningkatan *true positive* yang kompetitif, namun disertai jumlah *false positive* yang lebih besar, sehingga menandakan adanya trade-off antara sensitivitas dan ketepatan prediksi.



Gambar 6. Confusion Matrix Base Model

Evaluasi kinerja model dasar menggunakan kurva ROC–AUC menunjukkan bahwa ketiga algoritma memiliki kemampuan diskriminasi yang sangat baik. Berdasarkan Gambar 7, seluruh kurva ROC berada jauh di atas garis tebakan acak, dengan nilai AUC masing-masing sebesar 0,9838 untuk XGBoost, 0,9885 untuk LightGBM, dan 0,9850 untuk CatBoost. LightGBM memperoleh nilai AUC tertinggi, yang mengindikasikan kemampuan paling optimal dalam membedakan kelas positif dan negatif pada berbagai ambang keputusan, sementara XGBoost dan CatBoost menunjukkan performa yang kompetitif dan relatif seimbang.



Gambar 7. Kurva ROC-AUC Base Models

D. Hasil Stacking

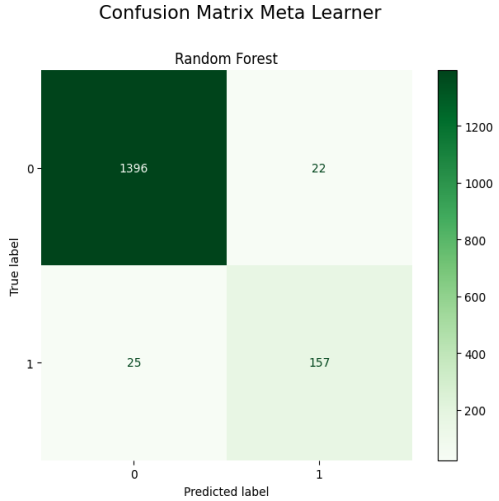
Tahap stacking merealisasikan pendekatan ensemble learning dengan mengkombinasikan prediksi probabilitas dari tiga model dasar, yaitu XGBoost, LightGBM, dan CatBoost, sebagai meta-features. Himpunan fitur ini kemudian digunakan untuk melatih Random Forest sebagai meta-learner (Level-1) guna menghasilkan keputusan klasifikasi akhir. Random Forest dikonfigurasi secara manual untuk menjaga stabilitas dan kemampuan generalisasi model. Hasil evaluasi kinerja model stacking yang disajikan pada Tabel 4.

TABEL 4
HASIL EVALUASI STACKING RF

Model	Accuracy	Precision	Recall	F1-Score
Stacking RF	0.9706	0.8771	0.8626	0.8698

Hasil evaluasi menunjukkan bahwa metode stacking memberikan performa terbaik, dengan akurasi mencapai 0.9706 dan melampaui model dasar terbaik. Nilai F1-score kelas positif juga meningkat menjadi 0.8698, yang menegaskan efektivitas integrasi model bertingkat dalam meningkatkan kinerja klasifikasi. Hal ini diperkuat oleh

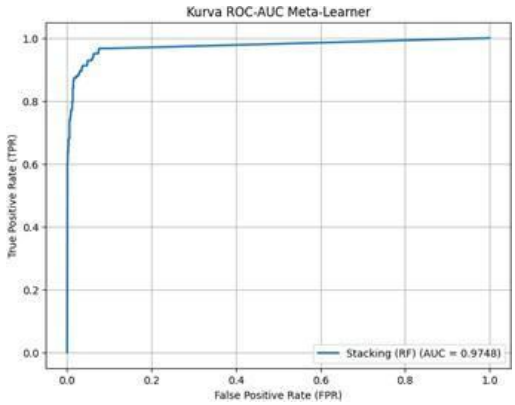
confusion matrix pada Gambar 8, yang menunjukkan dominasi prediksi benar (*true negative* dan *true positive*) serta jumlah kesalahan klasifikasi yang relatif rendah pada kedua kelas.



Gambar 8. Confusion Matrix Stacking RF

Distribusi nilai True Positive, True Negative, False Positive, dan False Negatif menunjukkan bahwa seluruh proses prediksi berhasil dieksekusi tanpa kegagalan komputasi. Selain itu, kesesuaian antara Confusion Matrix dengan nilai Accuracy, Precision, Recall, dan F1-Score memperlihatkan konsistensi hasil evaluasi yang dihasilkan oleh implementasi model.

Kurva ROC-AUC pada Gambar 9 menunjukkan skor 0.9748. Bentuk kurva yang melengkung tajam ke sudut kiri atas menandakan bahwa model *Stacking* memiliki kemampuan diskriminasi yang sangat baik dalam membedakan air aman dan tidak aman di berbagai ambang batas. Meskipun skor AUC terlihat sedikit berbeda dari variasi model dasar, stabilitas keputusan yang ditunjukkan oleh metrik klasifikasi (Precision/Recall) menegaskan bahwa *Stacking* adalah model yang paling reliabel untuk keputusan akhir.



Gambar 9. Kurva ROC-AUC Stacking RF

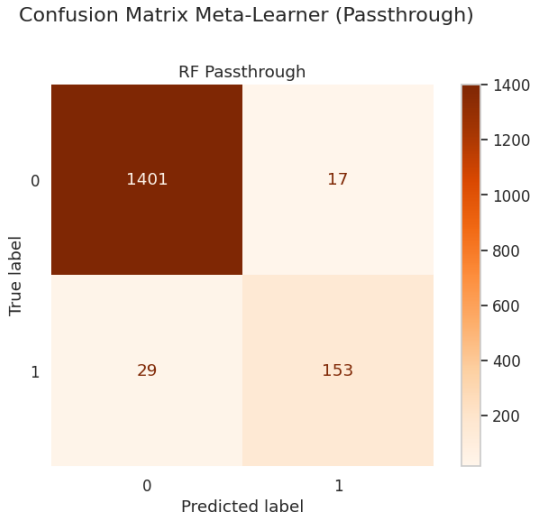
E. Hasil Stacking dengan Penambahan Fitur Original

Sebagai pengembangan arsitektur stacking, diterapkan teknik *feature propagation* dengan mengaktifkan parameter *passthrough*, yang menggabungkan probabilitas prediksi model dasar dan fitur asli dataset sebagai masukan bagi Random Forest selaku *meta-learner*. Pendekatan ini memungkinkan pemanfaatan informasi yang lebih komprehensif dan, sebagaimana ditunjukkan pada Tabel 5, menghasilkan peningkatan kinerja yang konsisten dibandingkan stacking tanpa *passthrough*.

TABEL 5
HASIL EVALUASI STACKING RF DENGAN FITUR ORIGINAL

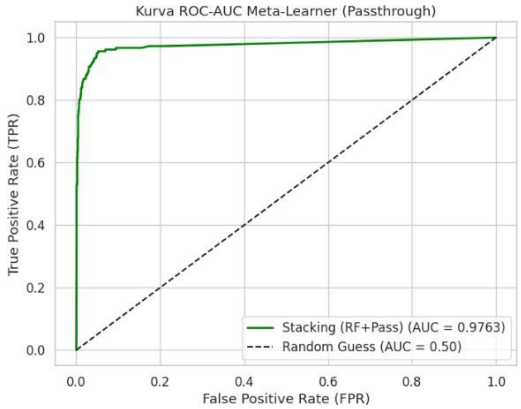
Model	Accuracy	Precision	Recall	F1-Score
Stacking RF + Passthrough	0.9712	0.9000	0.8407	0.8693

Hasil evaluasi menunjukkan bahwa *Stacking Random Forest dengan passthrough* memberikan performa terbaik, dengan akurasi 0.9712, precision 0.9000, dan F1-score 0.8693. Keunggulan ini diperkuat oleh *confusion matrix* pada Gambar 10, yang menunjukkan penurunan kesalahan klasifikasi menjadi 17 sampel serta peningkatan *true negative* hingga 1.401, sehingga model lebih selektif dan efektif dalam meminimalkan risiko alarm palsu.



Gambar 10. Confusion Matrix RF dengan Original Fitur

Kurva ROC–AUC pada Gambar 10 menunjukkan peningkatan nilai AUC menjadi 0.9763, lebih tinggi dibandingkan stacking tanpa *passthrough*. Bentuk kurva yang mendekati sudut kiri atas mengindikasikan peningkatan kemampuan separasi kelas, sehingga menegaskan bahwa integrasi fitur asli melalui mekanisme *passthrough* menghasilkan konfigurasi stacking yang lebih optimal untuk klasifikasi kualitas air.



Gambar 11. Kurva ROC-AUC Stacking RF dengan Original Fitur

F. Hasil Hyperparameter Tuning

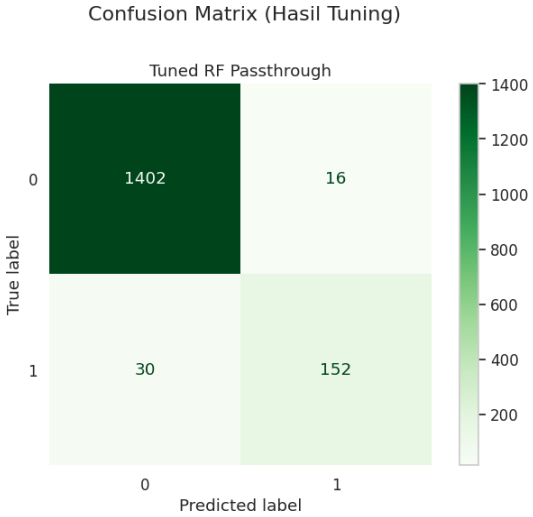
Berdasarkan hasil evaluasi sebelumnya, model *Stacking Random Forest dengan fitur original* dipilih sebagai arsitektur terbaik dan selanjutnya dioptimasi melalui proses *hyperparameter tuning*. Penyetelan parameter dilakukan menggunakan metode *Randomized Search CV* karena efisiensinya dalam mengeksplorasi ruang parameter yang luas. Optimasi difokuskan pada parameter *meta-learner* untuk memperoleh konfigurasi yang mampu menyeimbangkan bias dan varians. Hasil konfigurasi optimal yang diperoleh dari proses tuning disajikan pada Tabel 6.

TABEL 6
HASIL EVALUASI HYPERPARAMETER TUNING

Model	Accuracy	Precision	Recall	F1-Score
Stacking RF Pass Tuning	0.9712	0.9048	0.8352	0.8686

Hasil evaluasi pasca *hyperparameter tuning* menunjukkan bahwa model mempertahankan akurasi sebesar 0.9712, dengan peningkatan nilai precision menjadi 0.9048. Perubahan distribusi kesalahan terlihat pada *confusion matrix* pada Gambar 12, di mana jumlah *false positive* menurun dari 17 menjadi 16 sampel. Penurunan ini menunjukkan bahwa meta-learner mampu memanfaatkan informasi prediksi dari XGBoost, LightGBM, dan CatBoost

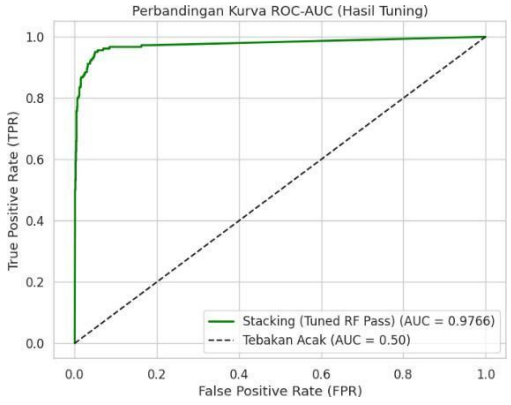
untuk memperbaiki keputusan klasifikasi terhadap sampel yang sebelumnya masih salah diprediksi sebagai kelas aman. Dalam konteks klasifikasi kualitas air, False Positive merupakan jenis kesalahan yang paling kritis, yaitu kondisi ketika air yang sebenarnya tidak layak konsumsi diprediksi sebagai air yang layak. Kesalahan seperti ini berpotensi menyebabkan air yang masih mengandung kontaminan kimia maupun biologis direkomendasikan untuk digunakan sebagai air minum. Dampaknya tidak hanya menurunkan reliabilitas sistem klasifikasi, tetapi juga dapat meningkatkan risiko gangguan kesehatan apabila hasil prediksi dijadikan dasar pengambilan keputusan oleh pengguna maupun instansi pengelola kualitas air. Meskipun *false negative* sedikit meningkat, penurunan kesalahan prediksi kritis ini menunjukkan bahwa proses tuning membuat model lebih selektif dan efektif dalam meminimalkan risiko kesalahan klasifikasi.



Gambar 12. Confusion Matrix Hyperparameter Tuning

Kurva ROC–AUC pada Gambar 13 menunjukkan nilai AUC sebesar 0.9766, sedikit lebih tinggi dibandingkan model sebelum tuning. Peningkatan ini mengindikasikan perbaikan kemampuan separasi kelas pada berbagai ambang keputusan, sementara bentuk kurva yang konsisten mendekati sudut kiri atas menegaskan bahwa konfigurasi hiperparameter hasil *Randomized Search* bersifat stabil dan robust.

Seluruh hasil evaluasi diperoleh melalui proses eksekusi model pada data uji menggunakan konfigurasi hyperparameter terbaik hasil *RandomizedSearchCV*. Konsistensi antara nilai metrik klasifikasi, Confusion Matrix, dan kurva ROC-AUC menunjukkan bahwa model dapat dijalankan secara stabil pada lingkungan komputasi yang digunakan. Dengan demikian, implementasi algoritma tidak hanya memberikan peningkatan performa, tetapi juga memenuhi aspek validasi komputasi yang menunjukkan bahwa model telah berhasil diimplementasikan dan dievaluasi secara menyeluruh.



Gambar 13. Kurva ROC-AUC Hyperparameter Tuning

G. Komparasi Performa Base Learner dan Ensemble Stacking

Berdasarkan hasil evaluasi base learner, yaitu algoritma berbasis gradient boosting memberikan performa klasifikasi yang tinggi dengan nilai accuracy di atas 95% seperti yang disajikan pada tabel 7. Meskipun demikian, masing-masing model masih memiliki karakteristik kesalahan prediksi yang berbeda sehingga belum mampu memberikan performa optimal secara konsisten. Integrasi ketiga model melalui pendekatan Ensemble Stacking RF menghasilkan peningkatan accuracy menjadi 97,06%, lebih tinggi dibandingkan seluruh base learner. Peningkatan ini

menunjukkan bahwa meta-learner Random Forest mampu mengkombinasikan kekuatan masing-masing algoritma sehingga kesalahan prediksi yang masih muncul pada model tunggal dapat dikoreksi melalui proses pembelajaran tingkat kedua (*second-level learning*). Selain meningkatkan accuracy, model Stacking juga menghasilkan peningkatan F1-Score menjadi 86,98%, yang menunjukkan keseimbangan yang lebih baik antara precision dan recall.

Ketika mekanisme passthrough diterapkan, accuracy meningkat menjadi 97,12%, sedangkan precision meningkat menjadi 90,00%. Hal ini menunjukkan bahwa penambahan fitur asli bersama keluaran base learner memberikan informasi tambahan yang membantu meta-learner dalam membedakan sampel dengan karakteristik yang saling berdekatan. Setelah dilakukan hyperparameter tuning, accuracy tetap dipertahankan pada 97,12%, sementara precision kembali meningkat menjadi 90,48%. Walaupun peningkatan accuracy hanya sebesar 0,43% dibandingkan model tunggal terbaik (LightGBM), peningkatan precision sebesar 4,84% dan penurunan jumlah False Positive menunjukkan bahwa model yang diusulkan memiliki kemampuan generalisasi yang lebih baik serta lebih aman diterapkan pada sistem klasifikasi kualitas air karena mampu meminimalkan risiko kesalahan dalam mengidentifikasi air yang sebenarnya tidak layak konsumsi sebagai air yang aman.

TABEL 7
PERBANDINGAN PERFORMA MODEL

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
XGBoost	96.25	85.06	81.32	83.15
CatBoost	95.87	79.29	86.26	82.63
LightGBM	**96.69**	85.64	85.16	85.40
Stacking RF	97.06	87.71	**86.26**	**86.98**
Stacking RF + Passthrough	97.12	90.00	84.07	86.93
Proposed (Passthrough + Tuning)	**97.12**	**90.48**	83.52	86.86

IV. SIMPULAN

Hasil penelitian menunjukkan bahwa penerapan metode ensemble stacking memberikan kinerja klasifikasi kualitas air yang paling optimal dibandingkan seluruh model yang diuji. Stacking yang mengombinasikan XGBoost, LightGBM, dan CatBoost dengan Random Forest sebagai meta-learner mampu meningkatkan akurasi menjadi 97,06%, melampaui performa masing-masing model dasar. Penambahan fitur original melalui mekanisme passthrough terbukti memperkuat kemampuan model dengan menaikkan akurasi hingga 97,12% dan meningkatkan precision secara signifikan di angka 90,00%, sehingga kesalahan prediksi terutama false positive dapat ditekan. Proses hyperparameter tuning selanjutnya menjaga stabilitas akurasi tetap di angka 97,12% sekaligus menghasilkan precision tertinggi 90,48%, yang menandakan model semakin selektif dan andal dalam menentukan status kualitas air. Temuan ini menegaskan bahwa strategi stacking yang dioptimasi efektif menghasilkan model klasifikasi yang robust dan berpotensi diterapkan sebagai komponen pendukung dalam sistem pemantauan kualitas air, baik pada instansi pengelola sumber daya air maupun laboratorium pengujian lingkungan. Kemampuan model dalam menekan kesalahan False Positive membantu mengurangi risiko air yang sebenarnya tidak layak konsumsi diklasifikasikan sebagai air yang aman, sehingga dapat mendukung proses pengambilan keputusan yang lebih akurat dan meningkatkan aspek keselamatan masyarakat. Sebagai arah pengembangan penelitian selanjutnya, disarankan untuk mengeksplorasi kombinasi *base learner* yang lebih beragam atau menerapkan *meta-learner* alternatif, serta memperluas cakupan dataset guna meningkatkan generalisasi dan robustnes model.

UCAPAN TERIMAKASIH

Ucapan terima kasih disampaikan kepada Universitas Amikom Yogyakarta atas dukungan akademik, lingkungan ilmiah, serta fasilitas yang diberikan sehingga penelitian ini dapat terlaksana dengan baik. Terima kasih juga disampaikan kepada dosen pembimbing atas bimbingan, arahan, dan masukan yang konstruktif selama proses penelitian dan penulisan, sehingga penelitian ini dapat diselesaikan secara sistematis dan sesuai dengan kaidah ilmiah.

DAFTAR PUSTAKA

[1] N. Akmal, M. Faez, A. Jalal, K. Chowdhury, and N. Hafizah, "Water pollution and the assessment of water quality parameters : a review," *Desalin. Water Treat.*, vol. 294, pp. 79–88, 2023, doi: 10.5004/dwt.2023.29433.

[2] I. Juwana, S. V. Lazuardi, and H. Herdiansyah, "Investigating the Factor of Water Consumption Regarding the Impact and Implementation of Water Governance in Urban Areas," vol. 7, no. 1, pp. 55–64, 2024.

[3] W. P. Anggraini, D. I. Kusumastuti, E. P. Wahono, T. Sipil, and U. Lampung, "Sistem Informasi Kualitas Air Sungai di Wilayah Sungai Seputih," vol. 9, no. 4, pp. 964–978, 2024.

[4] H. Zhang *et al.*, "Natural and anthropogenic imprints on seasonal river water quality trends across China," 2025, doi: 10.1038/s41545-025-00481-3.

[5] M. Zhu *et al.*, "A review of the application of machine learning in water quality evaluation," *Eco-Environment Heal.*, vol. 1, no. 2, pp. 107–116, Jun. 2022, doi: 10.1016/j.eehl.2022.06.001.

[6] U. Rohima Zalti, D. Rose Darmakusuma, M. Ridwansyah, and E. Ismanto, "Analisis dan Prediksi Kelayakan Air Minum Menggunakan Algoritma Random Forest," *J. FASILKOM*, vol. 15, no. 2, pp. 312–317, Aug. 2025, doi: 10.37859/jf.v15i2.9906.

[7] T. Z. Jasman, M. A. Fadhlullah, A. L. Pratama, and R. Rismayani, "Analisis Algoritma Gradient Boosting, Adaboost dan Catboost dalam Klasifikasi Kualitas Air," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 2, Aug. 2022, doi: 10.28932/jutisi.v8i2.4906.

[8] M. Dava Maulana, A. Id Hadiana, and F. Rakhmat Umbara, "Algoritma Xgboost Untuk Klasifikasi Kualitas Air Minum," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 5, pp. 3251–3256, 2024, doi: 10.36040/jati.v7i5.7308.

[9] Prasetya Widiharso, Siti Sendari, Anik Nur Handayani, and Nastiti Susetyo Fanani Putri, "Performa Metode Klasifikasi Tunggal dan Ensemble Model dalam Identifikasi Baku Mutu Air," *Infotekmesin*, vol. 13, no. 2, pp. 206–211, Jul. 2022, doi: 10.35970/infotekmesin.v13i2.1529.

[10] B. M. Karomah, "PENERAPAN METODE STACKING DALAM MENGKLASIFIKASIKAN PENDERITA PENYAKIT DIABETES," *J. Publ. Ilmu Komput. dan Multimed.*, vol. 1, no. 3, pp. 188–194, Jan. 1970, doi: 10.55606/jupikom.v1i3.522.

[11] E. Prasetyo and K. Nugroho, "Optimasi Klasifikasi Data Stunting Melalui Ensemble Learning pada Label Multiclass dengan Imbalance Data," *Techno.Com*, vol. 23, no. 1, pp. 1–10, Feb. 2024, doi: 10.62411/tc.v23i1.9779.

[12] M. Hermansyah, A. Saikhu, and B. Amaliah, "Pemodelan data radiosonde menggunakan stacking ensemble untuk klasifikasi hujan," vol. 10, no. 2, pp. 1678–1687, 2025.

[13] M. Kumar, S. Singhal, S. Shekhar, B. Sharma, and G. Srivastava, "Optimized Stacking Ensemble Learning Model for Breast Cancer Detection and Classification Using Machine Learning," *Sustainability*, vol. 14, no. 21, p. 13998, Oct. 2022, doi: 10.3390/su142113998.

[14] A. S. Alfath and A. K. Wardhana, "Hypertension Risk Prediction Using Stacking Ensemble of CatBoost , XGBoost , and LightGBM : A Machine Learning Approach," vol. 9, no. 6, pp. 3146–3156, 2025.

[15] F. M. Model, "Stacking Ensemble Learning : Combining XGBoost , LightGBM , CatBoost , and AdaBoost with Random," 2025.

[16] A. A. Saputra, B. N. Sari, and C. Rozikin, "Penerapan Algoritma Extreme Gradient Boosting (Xgboost) Untuk Analisis Risiko Kredit," vol. 05, no. 01, pp. 53–64, 2022.

[17] F. I. Kurniadi and P. D. Larasati, "Light Gradient Boosting Machine untuk Deteksi Penyakit Stroke," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 6, no. 1, pp. 67–72, 2022, doi: 10.47970/siskom-kb.v6i1.328.

[18] A. Fahmi, L. Ptr, M. M. Siregar, and I. Daniel, "Journal of Computer Networks , Architecture and High Performance Computing Analysis of Gradient Boosting , XGBoost , and CatBoost on Mobile Phone Classification Journal of Computer Networks , Architecture and High Performance Computing," vol. 6, no. 2, pp. 661–670, 2024.

[19] D. Feby, "Serba Serbi Machine Learning Model Random Forest." [Online]. Available: https://dqlab.id/serba-serbi-machine-learning-model-random-forest

[20] A. D. Rachmatsyah, T. Sugihartono, and K. Irfan, "PERBANDINGAN TEKNIK OPTIMASI GRID SEARCH DAN RANDOMIZED SEARCH DALAM MENINGKATKAN AKURASI METODE KLASIFIKASI SVM PADA SENTIMEN ULASAN PENGGUNA APLIKASI JKN MOBILE," *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 8, no. 1, pp. 13–22, Dec. 2024, doi: 10.36080/skanika.v8i1.3328.

[21] S. Q. Sultan, N. Javaid, and N. Alrajeh, "Machine Learning-Based Stacking Ensemble Model for Prediction of Heart Disease with Explainable AI and K-Fold Cross-Validation : A Symmetric Approach," no. Cvd, pp. 1–26, 2025.

[22] H. S. Silva *et al.*, "The Impact of Feature Scaling In Machine Learning : Effects on Regression and Classification Tasks," 2025, doi: 10.1109/ACCESS.2025.3635541.

[23] C. Paramita, C. S. Simbolon, A. S. Pamungkas, J. M. Triono, P. W. Utomo, and E. R. Subhiyakto, "Jurnal Informatika : Jurnal pengembangan IT Analisis Pengaruh SMOTE terhadap Kinerja Model KNN untuk Prediksi Risiko Stroke," vol. 10, no. 4, pp. 978–988, 2025, doi: 10.30591/jpit.v10i4.8809.

[24] K. S. Hasanah, Uswatun Agus Mohamad Soleh, "Effect of Random Under sampling , Oversampling , and SMOTE on the Performance of Cardiovascular Disease Prediction Models terhadap Kinerja Model Prediksi Penyakit Kardiovaskular," vol. 21, no. 1, pp. 88–102, 2024, doi: 10.20956/j.v21i1.35552.

[25] G. Erutjahjo and A. Supriyanto, "Jurnal Informatika : Jurnal pengembangan IT Prediksi Tinggi Gelombang Laut di Perairan Semarang – Demak dengan Menggunakan Random Forest dan XGBoost," vol. 10, no. 4, pp. 869–881, 2025, doi: 10.30591/jpit.v10i4.9315.