



Implementation of LightGBM Algorithm for Determining Aquarium Water Conditions Based on Water Quality Parameters

Penerapan Algoritma LightGBM untuk Penentuan Kondisi Air Akuarium Berdasarkan Parameter Kualitas Air

Mutiara Qolbu¹, Ahmad Taqwa^{2*}, Mohammad Fadhli³

^{1,2,3}Program Studi Teknik Telekomunikasi, Politeknik Negeri Sriwijaya, Indonesia

E-Mail: ¹mutiaraqllbb@gmail.com, ²taqwa@polsri.ac.id, ³mohammad.fadhli@polsri.ac.id

Received Jun 13th 2026; Revised Jul 16th 2026; Accepted Jul 27th 2026; Available Online Aug 16th 2026

Corresponding Author: Ahmad Taqwa

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

Water quality is crucial for the health and survival of aquarium fish. Changes in pH, temperature, total dissolved solids (TDS), and ammonia levels can create unsuitable water conditions for fish. Conventional monitoring methods are limited in their ability to simultaneously analyze the relationships among these parameters. This study aims to use the LightGBM algorithm to classify aquarium water quality based on pH, temperature, TDS, and ammonia. The dataset comprises 710 water-quality sensor measurements. These data underwent several stages: preprocessing, labeling, segmentation, hyperparameter tuning, and model training. Test results showed that the best LightGBM configuration achieved 99.30% accuracy, 98.61% precision, 99.28% recall, and 98.92% F1-score. Evaluation using a confusion matrix showed that of 142 test data sets, only one misclassification occurred, with all good and bad categories correctly predicted. These results demonstrate that LightGBM can classify aquarium water quality with high accuracy.

Keywords: Ammonia, Classification, Hyperparameter, LightGBM, Machine Learning, pH, Water Quality

Abstrak

Kualitas air sangat penting bagi kesehatan dan kelangsungan hidup ikan di akuarium. Jika parameter seperti pH, suhu, *total dissolved solids* (TDS), dan amonia berubah, kondisi air dapat menjadi tidak cocok bagi ikan. Masalahnya, metode pemantauan biasa masih terbatas dalam menganalisis hubungan antara parameter-parameter tersebut secara simultan. Penelitian ini bertujuan untuk menggunakan algoritma LightGBM untuk mengklasifikasikan kualitas air akuarium berdasarkan pH, suhu, TDS, dan amonia. Dataset yang digunakan terdiri dari 710 data pengukuran sensor kualitas air. Data-data ini melalui beberapa tahap, yaitu preprocessing, pelabelan data, pembagian data, penyesuaian hyperparameter, dan pelatihan model. Hasil pengujian menunjukkan bahwa konfigurasi LightGBM terbaik memiliki *accuracy* 99,30%, *precision* 98,61%, *recall* 99,28%, dan *F1-score* 98,92%. Evaluasi menggunakan confusion matrix menunjukkan bahwa dari 142 data pengujian hanya terdapat satu kesalahan klasifikasi, dengan seluruh data kategori baik dan buruk berhasil diprediksi dengan tepat. Hasil tersebut menunjukkan bahwa LightGBM dapat mengklasifikasikan kualitas air akuarium dengan akurasi yang tinggi.

Kata Kunci: Amonia, Hyperparameter, Klasifikasi, Kualitas Air, LightGBM, Machine Learning, pH

1. PENDAHULUAN

Keadaan suatu ekosistem perairan tidak terlepas dari perubahan kualitas air yang dapat berdampak pada kesehatan, pertumbuhan, perilaku, dan kelangsungan hidup ikan. Hal ini karena perubahan parameter fisika dan kimia, seperti suhu, pH, total dissolved solids (TDS), oksigen terlarut, dan amonia, dapat memengaruhi proses fisiologis, respons perilaku, serta kemampuan adaptasi ikan terhadap lingkungannya [1]. Selain itu, TDS yang tinggi dapat mengganggu proses pertumbuhan dan kelangsungan hidup ikan. Di sisi lain, peningkatan konsentrasi amonia dalam air dapat merusak fungsi insang, mengurangi ketahanan terhadap penyakit, dan menurunkan kinerja pertumbuhan [2]. Demikian pula, pH yang tidak sesuai dapat menyebabkan stres fisiologis yang parah dan meningkatkan risiko kematian ikan. Hasil penelitian sebelumnya mengungkapkan bahwa nilai pH yang ekstrem dapat menyebabkan kematian ikan hingga 50% dalam situasi tertentu [2]. Temuan ini

menyoroti pentingnya menjaga kualitas air yang memadai untuk mendukung kehidupan dan kesejahteraan ikan.

Kualitas air juga berperan penting dalam produktivitas dan pengelolaan budidaya perikanan. Produksi ikan sangat dipengaruhi oleh kondisi lingkungan, dan penurunan kualitas air dapat memberikan dampak negatif terhadap organisme akuatik, seperti munculnya penyakit, penurunan laju pertumbuhan, hingga kematian ikan [3], [4]. Maka dari itu, pemantauan berkelanjutan terhadap parameter kualitas air seperti suhu, pH, oksigen terlarut, kekeruhan, *total dissolved solids* (TDS), dan konsentrasi amonia menjadi bagian penting dalam pengelolaan budidaya ikan [3]. Namun, metode penilaian kualitas air konvensional umumnya masih bergantung pada ambang batas yang telah ditentukan dan perhitungan berbasis parameter tertentu, sehingga belum mampu menggambarkan interaksi kompleks antarparameter kualitas air secara menyeluruh [5]. Selain itu, pendekatan tersebut sering melibatkan pembobotan yang bersifat subjektif sehingga kontribusi masing-masing parameter terhadap kondisi kualitas air secara keseluruhan belum dapat direpresentasikan secara akurat [6]. Keterbatasan tersebut menjadi permasalahan penting dalam pemantauan kualitas air akuarium karena perubahan kondisi air dapat terjadi dengan cepat akibat aktivitas pemberian pakan, akumulasi sisa metabolisme, serta perubahan kondisi lingkungan. Perubahan tersebut sering kali sulit dikenali melalui pengamatan manual, sehingga keterlambatan dalam penanganannya dapat meningkatkan risiko stres hingga kematian ikan. Oleh karena itu, diperlukan metode yang mampu menganalisis berbagai parameter kualitas air secara bersamaan untuk menghasilkan klasifikasi kondisi air yang lebih akurat.

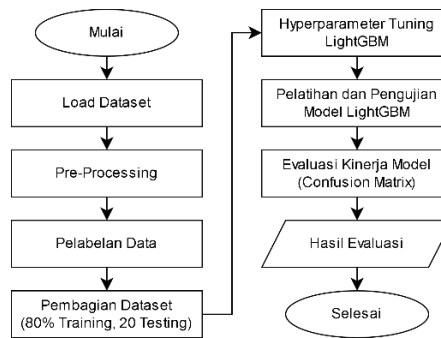
Kemajuan terkini dalam sektor *machine learning* telah menawarkan solusi yang menjanjikan untuk pemantauan lingkungan dan penilaian kualitas air. Model *machine learning* mampu memproses kumpulan data berskala besar dan multidimensi, mengidentifikasi pola tersembunyi, serta mempelajari hubungan nonlinier yang kompleks antar variabel [5]. Studi sebelumnya menunjukkan bahwa algoritma machine learning mampu digunakan untuk mengklasifikasikan dan memprediksi kondisi kualitas air berdasarkan berbagai indikator lingkungan [6]. Selain itu, penerapan teknik pembelajaran mesin telah menunjukkan performa yang baik dalam analisis kualitas air serta membantu proses pengambilan keputusan berbasis data [7]. Beberapa penelitian lain juga membuktikan bahwa machine learning mampu mengenali hubungan kompleks antarparameter kualitas air, seperti prediksi parameter akuakultur [8], klasifikasi kondisi perairan berbasis data sensor [9], serta pemantauan kualitas air secara *real-time* melalui integrasi IoT [10] [11]. Penelitian lainnya menunjukkan bahwa pendekatan berbasis machine learning dapat digunakan untuk mendeteksi perubahan kualitas air yang berpotensi memengaruhi kesehatan ikan [12]. Hasil tersebut menunjukkan bahwa machine learning memiliki kemampuan mengolah data lingkungan yang kompleks dan menghasilkan informasi kualitas air secara otomatis.

Di antara berbagai pendekatan *machine learning*, *Light Gradient Boosting Machine* (LightGBM) telah muncul sebagai salah satu metode yang ampuh untuk penilaian kualitas air. LightGBM dirancang untuk memproses dataset besar secara efisien dengan waktu pelatihan yang cepat, penggunaan memori yang rendah, serta kemampuan yang baik dalam mempelajari hubungan kompleks antarvariabel [13]. Penelitian sebelumnya membuktikan LightGBM mampu mencapai kinerja tertinggi di antara beberapa model *machine learning*, dengan akurasi sebesar 99,65% dalam klasifikasi kualitas air [7], [14]. Demikian pula, penelitian lain menunjukkan bahwa LightGBM mencapai akurasi sebesar 95,1% dalam penilaian kualitas air sekaligus secara efektif mengurangi kesalahan klasifikasi [13]. Lebih lanjut, pada studi klasifikasi kualitas air berskala besar yang melibatkan lebih dari 2,5 juta data, LightGBM berhasil mencapai akurasi sebesar 98% [15]. Temuan ini menunjukkan bahwa LightGBM memiliki potensi yang kuat untuk mengklasifikasikan kondisi kualitas air secara akurat.

Meskipun berbagai penelitian telah menunjukkan kemampuan machine learning dalam analisis kualitas air, sebagian besar studi sebelumnya masih berfokus pada air permukaan, sistem hidrologi, atau lingkungan budidaya perairan skala besar [5]–[15]. Penelitian terdahulu umumnya memanfaatkan *machine learning* untuk memprediksi nilai parameter tertentu, seperti pH, oksigen terlarut, atau amonia, sehingga belum secara langsung memberikan informasi yang menyeluruh tentang kondisi kualitas air. Selain itu, penerapan *machine learning* pada lingkungan akuarium masih relatif terbatas, padahal perubahan kualitas air pada akuarium dapat terjadi akibat aktivitas pemeliharaan ikan serta kondisi lingkungan yang dinamis. Oleh karena itu, penelitian ini menerapkan algoritma LightGBM untuk mengklasifikasikan kondisi kualitas air akuarium berdasarkan kombinasi parameter pH, suhu, TDS, dan amonia ke dalam kategori baik, waspada, dan buruk. Pengembangan model dilakukan melalui pengujian beberapa konfigurasi *hyperparameter* untuk memperoleh model yang mampu mengenali pola kualitas air secara optimal.

2. BAHAN DAN METODE

Metode yang diterapkan dalam studi ini adalah pendekatan klasifikasi menggunakan algoritma LightGBM untuk mengkategorikan kondisi kualitas air akuarium berdasarkan parameter yang telah ditentukan. Proses penelitian ini mencakup pengumpulan dataset, prapemrosesan data, pelabelan data, pembagian data, penyesuaian *hyperparameter*, pelatihan model LightGBM, serta evaluasi model menggunakan *matriks konfusi*. Alur metodologi penelitian dapat dilihat pada Gambar 1.



Gambar 1. Metodologi Penelitian

2.1. Dataset

Dataset penelitian diperoleh melalui proses pengukuran sejumlah parameter kualitas air akuarium yang telah ditetapkan. Data ini diperoleh menggunakan beberapa sensor untuk memantau kualitas air. Pengumpulan data dilakukan secara berkala di akuarium yang berada di salah satu fasilitas penjualan ikan di Kota Palembang. Setiap data ini berisi nilai parameter kualitas air yang nantinya digunakan sebagai variabel masukan dalam proses klasifikasi. Untuk memahami atribut-atribut yang digunakan, kita dapat melihat deskripsi atribut pada Tabel 1.

Tabel 1. Deskripsi Atribut Dataset

Atribut	Keterangan	Satuan
pH	Derajat keasaman atau kebasaan air	-
Suhu	Temperatur air akuarium	°C
Amonia	Konsentrasi amonia dalam air	ppm
TDS	Total padatan terlarut dalam air	Ppm

2.2. Prapemrosesan Data

Prapemrosesan data (*data preprocessing*) merupakan proses persiapan yang bertujuan untuk meningkatkan kualitas data sebelum diterapkan pada algoritma *machine learning*. Dataset hasil pengukuran pada kondisi nyata sering kali belum memiliki kualitas yang sempurna, karena masih dapat ditemukan permasalahan berupa nilai kosong, ketidaksesuaian data, maupun gangguan lain sehingga berpotensi memengaruhi proses analisis. Maka dari itu, proses prapemrosesan diperlukan untuk membersihkan dan menyempurnakan data agar lebih reliabel serta sesuai untuk digunakan dalam pengembangan model *machine learning* [16].

Proses prapemrosesan dilakukan untuk menangani permasalahan pada dataset agar kualitas data meningkat dan model mampu memperoleh pola yang lebih representatif. Secara umum, prapemrosesan dapat mencakup berbagai tahapan seperti membersihkan data, menggabungkan data, transformasi data, reduksi dimensi, dan pelabelan data sesuai dengan kebutuhan penelitian [17]. Melalui tahapan prapemrosesan yang sesuai, informasi pada dataset dapat diperbaiki sehingga kinerja dan kemampuan model untuk generalisasi menjadi lebih optimal [16], [17].

2.3. Pelabelan Data

Tahap pelabelan data dilakukan dengan memberikan kategori atau kelas pada setiap data berdasarkan karakteristik tertentu yang dimilikinya. Tahap ini berfungsi untuk mengumpulkan informasi target yang dapat dimanfaatkan dalam proses pelatihan model klasifikasi. Melalui proses pelabelan, setiap data dapat dikelompokkan ke dalam kategori yang sesuai sehingga membantu model *machine learning* untuk lebih mudah mengidentifikasi pola dan hubungan di antara variabel [18].

Dalam penelitian ini, proses pelabelan diterapkan pada *dataset* kualitas air yang terdiri dari hasil pengukuran parameter pH, suhu, amonia, dan TDS. Setiap parameter terlebih dahulu dievaluasi berdasarkan kondisi yang telah ditentukan, kemudian hasil evaluasi tersebut digunakan untuk memberikan label kualitas air. Data berikutnya diklasifikasikan ke dalam tiga kategori, yaitu baik, waspada, dan buruk, yang digunakan sebagai kelas target dalam proses klasifikasi menggunakan algoritma LightGBM.

2.4. Pembagian Data

Data yang telah melalui proses pelabelan kemudian dibagi menjadi dua kelompok, yaitu data yang digunakan pada tahap pembelajaran model dan data yang digunakan untuk menguji performa model. Tujuan yang ingin dicapai adalah menggunakan data latih untuk menciptakan model klasifikasi yang baik, kemudian menguji seberapa baik model tersebut bekerja dengan data yang belum pernah digunakan sebelumnya.

Pemisahan data ini sangat penting dalam *machine learning* sebab, hasilnya akan sangat berpengaruh pada kinerja model untuk menggeneralisasi data di luar proses pelatihan [19]. Pemisahan data pelatihan dan data pengujian dilakukan agar proses evaluasi kinerja model dapat dilakukan secara lebih objektif [20]. Kinerja model pun dapat diukur secara akurat. Tahap ini penting untuk mengetahui seberapa efektif performa model dalam melakukan klasifikasi. Dengan cara ini, kita bisa meningkatkan kualitas model dan memastikan model tersebut siap digunakan dalam situasi nyata.

Pada penelitian ini digunakan metode *train-test split* dengan perbandingan 80:20, yaitu sebagian besar data sebesar 80% dialokasikan sebagai data pelatihan dan 20% lainnya sebagai data pengujian. Pemilihan perbandingan 80:20 ini dilakukan karena penelitian sebelumnya menunjukkan bahwa dengan menggunakan lebih banyak data untuk pelatihan, model dapat lebih baik dalam memahami pola data [21]. Selain itu, perbandingan 80:20 juga memberikan proporsi yang seimbang antara proses pelatihan dan pengujian model [21]. Proses pembagian data dilakukan secara acak, tetapi tetap memastikan rasio masing-masing kelas dalam data tetap sama, sehingga sebaran kelas pada data pelatihan dan data pengujian tetap seimbang.

2.5. Machine Learning

Machine learning adalah salah satu pendekatan dalam kecerdasan buatan (*Artificial Intelligence*) yang memungkinkan sistem komputer melakukan proses pembelajaran dari data untuk membangun prediksi maupun keputusan secara otomatis tanpa instruksi pemrograman langsung [22], [23], [24]. Metode ini bekerja dengan menggunakan data historis sebagai sumber pembelajaran, sehingga model dapat memahami keterkaitan antarvariabel dan digunakan untuk menyelesaikan berbagai permasalahan klasifikasi maupun prediksi [1]. Dalam penerapan *machine learning*, *supervised learning* menjadi salah satu metode yang sering digunakan untuk membangun model berdasarkan data berlabel untuk membangun hubungan antara data masukan dan keluaran yang diharapkan [23], [25].

Perkembangan *machine learning* telah mendorong pemanfaatannya pada berbagai bidang, termasuk kesehatan, pertanian, industri, dan pemantauan lingkungan [22], [26]. Dalam bidang kualitas air, *machine learning* diterapkan guna menganalisis, mengklasifikasikan, dan memprediksi kualitas air berdasarkan data yang diperoleh dari sensor [27]. Kemampuannya untuk mengolah data secara otomatis serta mengidentifikasi pola yang rumit menjadikan *machine learning* sebagai teknologi yang ampuh untuk membantu memantau kualitas air dengan lebih cepat dan tepat. [26], [27]. Karena itu, penelitian ini menerapkan metode *machine learning* untuk mengklasifikasikan kondisi kualitas air akuarium berdasarkan parameter pH, suhu, amonia, dan TDS.

2.6. Light Gradient Boosting Machine (LightGBM)

Light Gradient Boosting Machine (LightGBM) adalah sebuah algoritma *machine learning* yang menggunakan metode *Gradient Boosting Decision Tree* (GBDT) [28]. Algoritma ini dikembangkan untuk membuat proses pelatihan pada data besar menjadi lebih efisien [29]. LightGBM menawarkan beberapa keunggulan, di antaranya waktu pelatihan yang lebih efisien, kebutuhan memori yang lebih kecil, serta kemampuan dalam menangani data dengan banyak fitur [30]. Kelebihan ini didapatkan berkat beberapa teknik optimasi yang digunakan, seperti *Gradient-based One-Side Sampling* (GOSS), *Exclusive Feature Bundling* (EFB), algoritma histogram, dan mekanisme pertumbuhan pohon *leaf-wise* [28], [31]. Teknik-teknik ini memungkinkan proses pembelajaran berlangsung lebih cepat tanpa mengurangi akurasi. Dengan demikian, LightGBM dapat menjadi pilihan yang baik untuk mengolah data besar dengan tetap mempertahankan tingkat akurasi yang tinggi.

Model LightGBM melakukan pembaruan prediksi secara iteratif sebagaimana ditunjukkan pada persamaan (1). Pada setiap iterasi, model menambahkan pohon keputusan baru untuk mengoptimalkan *loss function* dan meningkatkan hasil prediksi [32].

$$\hat{y}^{(t)} = \hat{y}^{(t-1)} + f_t(x_i) \quad (1)$$

Selain itu, LightGBM menggunakan fungsi regularisasi pada persamaan (2) untuk mengontrol kompleksitas model dan mengurangi risiko *overfitting* selama proses pelatihan [32].

$$\Omega(f) = \gamma T + \frac{\lambda}{2} \|\omega\|^2 \quad (2)$$

2.7. Hyperparameter Tuning

Hyperparameter tuning merupakan proses penentuan kombinasi parameter terbaik pada model *machine learning* guna mendapatkan performa yang maksimal. Berbeda dengan nilai parameter pada model yang otomatis ditemukan saat model dilatih, *hyperparameter* justru ditentukan sebelum proses pelatihan dimulai [33]. *Hyperparameter* ini nantinya akan mengatur bagaimana model belajar. Pemilihan *hyperparameter* yang

sesuai mampu membantu mengoptimalkan akurasi, performa generalisasi model terhadap data baru, serta menjaga keseimbangan antara kondisi *overfitting* dan *underfitting* [34], [35].

Tabel 2. Skenario *Hyperparameter* LightGBM

Skenario	Learning Rate	Max Depth	Num Leaves	N-Estimators
Skenario 1	0,005	2	3	30
Skenario 2	0,01	3	7	60
Skenario 3	0,02	4	8	100
Skenario 4	0,05	5	10	150
Skenario 5	0,10	6	20	200
Skenario 6	0,20	7	25	250

Pada penelitian ini, *hyperparameter tuning* diterapkan pada algoritma LightGBM dengan menguji enam skenario kombinasi nilai parameter yang meliputi *learning rate*, *max depth*, *num leaves*, dan *n_estimators*. Proses ini dilakukan melalui pelatihan dan evaluasi model secara berulang hingga diperoleh konfigurasi yang memberikan performa terbaik [36]. Pengaturan *hyperparameter* pada LightGBM menjadi tahapan penting karena nilai parameter yang digunakan dapat memengaruhi tingkat akurasi, efisiensi pelatihan, serta kemampuan model dalam membentuk pola klasifikasi dari data kualitas air [37], [38]. Perancangan skenario *hyperparameter* yang diterapkan dalam penelitian ini ditampilkan dalam Tabel 2.

2.8. Evaluasi Model (*Confusion Matrix*)

Evaluasi model dilakukan untuk mengetahui seberapa baik algoritma LightGBM bekerja dalam mengklasifikasikan kualitas air. Evaluasi dilakukan dengan membandingkan hasil klasifikasi model terhadap nilai aktual pada data menggunakan *confusion matrix*. Tahap berikutnya adalah mengukur performa model melalui perhitungan beberapa metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score* untuk melihat hasil kinerja model dalam melakukan klasifikasi [39].

a. *Accuracy*

Pengukuran *accuracy* bertujuan untuk mengetahui persentase prediksi yang sesuai antara keluaran model dan kondisi aktual pada dataset uji. Pengukuran *accuracy* dilakukan dengan melihat jumlah keluaran model yang sesuai dengan label aktual dibandingkan dengan total sampel pada data uji [32].

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (3)$$

b. *Precision*

Precision digunakan untuk melihat ketepatan model dalam mengidentifikasi suatu kelas [32]. Semakin besar nilai *precision*, semakin baik kemampuan model dalam menghasilkan prediksi yang sesuai pada kelas tersebut.

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

c. *Recall*

Recall digunakan untuk mengevaluasi seberapa efektif model dalam mengidentifikasi data yang cocok dengan kenyataan [32]. Nilai *recall* mencerminkan seberapa baik model bisa mendeteksi semua data pada kategori tertentu berdasarkan label yang sudah ditetapkan.

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (5)$$

d. *F1-Score*

F1-score digunakan untuk melihat keseimbangan antara *precision* dan *recall* [32]. Metrik ini memberikan nilai gabungan yang menunjukkan kemampuan model dalam melakukan prediksi secara akurat dan lengkap.

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \quad (6)$$

3. HASIL DAN PEMBAHASAN

3.1. Deskripsi Dataset

Dataset penelitian ini mencakup 710 pengamatan. Data ini dikumpulkan dari sebuah akuarium di fasilitas penjualan ikan di Palembang, Indonesia. Masing-masing pengamatan merupakan serangkaian pengukuran yang diambil oleh sensor kualitas air pada waktu tertentu. Dataset ini mencakup empat parameter

fisika dan kimia. Parameter-parameter tersebut adalah pH, suhu, kadar amonia, dan TDS. Keempat parameter ini digunakan sebagai data masukan untuk model klasifikasi.

Parameter yang diterapkan merupakan indikator kunci dalam menjelaskan keadaan air di dalam akuarium yang dapat dilihat pada Tabel 3. pH menunjukkan seberapa asam atau basa air tersebut, suhu mencerminkan keadaan termal di lingkungan air, amonia menunjukkan kadar senyawa nitrogen yang bersifat toksik, sedangkan TDS mencerminkan banyaknya zat terlarut yang ada dalam air. Perbedaan nilai pada parameter-parameter ini dapat memberikan informasi yang bermanfaat untuk mengenali variasi kondisi kualitas air, sehingga dapat digunakan dalam klasifikasi dengan bantuan *machine learning*.

Tabel 3. Contoh Dataset Parameter Air Aquarium

No	waktu	pH	suhu	amonia	tds
1	2026-05-25 11:41:53	7.10378	27.0	0.001970	81.48837
2	2026-05-25 11:51:02	7.33444	27.0	0.002420	77.96120
3	2026-05-21 19:04:09	6.30000	29.5	0.079039	265.00000
4	2026-05-24 20:07:22	4.88100	30.5	0.010390	36.44000
5	2026-05-25 11:51:23	6.63501	27.0	0.002190	75.59764

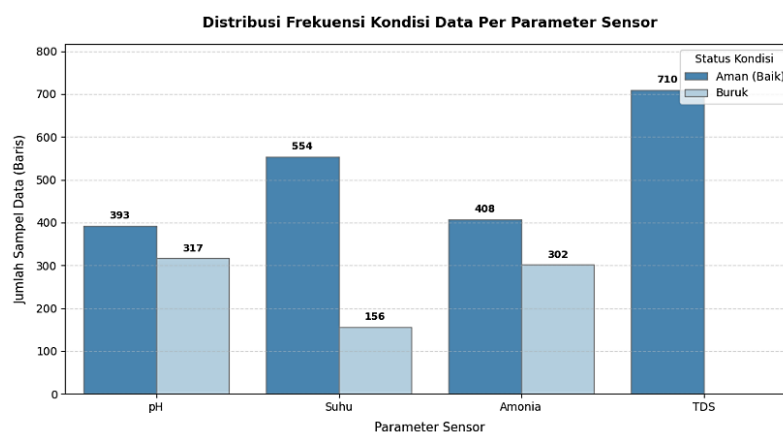
Dataset ini menjadi sumber utama untuk pengembangan dan evaluasi model. Karena dataset ini mengandung hasil pengukuran sensor yang sangat bervariasi, model klasifikasi dapat mempelajari pola-pola yang terkait dengan berbagai kondisi air akuarium berdasarkan karakteristik fisika-kimianya.

3.2. Pra-pemrosesan dan Pelabelan Data

Sebelum digunakan untuk klasifikasi, dataset harus melalui tahap prapemrosesan terlebih dahulu. Tujuannya adalah memastikan kualitas data yang digunakan untuk melatih model tetap terjaga. Pada tahap ini, data dengan nilai kosong akan dihapus dari dataset. Selain itu, data dengan nilai pH, suhu, atau TDS yang nol juga dihilangkan. Data-data tersebut dianggap tidak valid dan dapat memengaruhi proses pembelajaran model.

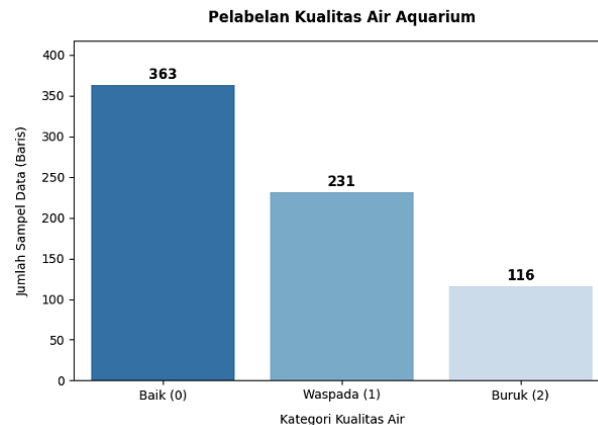
Pada tahap berikutnya, semua data yang diperoleh dari sensor diberi label tentang kualitas air. Label ini dibuat berdasarkan aturan yang sudah ditentukan sebelumnya. Tujuan pelabelan ini adalah mengubah data angka menjadi kategori kualitas air yang jelas, sehingga dapat digunakan sebagai acuan dalam algoritma LightGBM. Setiap parameter yang ada dievaluasi untuk menentukan apakah kondisinya aman atau tidak aman. Hasil evaluasi ini kemudian digunakan sebagai dasar untuk membuat label kualitas air yang sesuai.

Sebagai langkah awal, dilakukan analisis kondisi masing-masing parameter sensor yang terlihat pada Gambar 1. Dari hasil analisis ini, dapat dilihat bahwa semua data TDS berada dalam kondisi aman. Sementara itu, untuk parameter suhu, ditemukan bahwa sebagian besar data, yaitu 554 sampel, berada dalam kondisi aman. Namun, parameter pH dan amonia memiliki proporsi data yang tidak aman lebih tinggi dibandingkan dengan parameter lainnya. Dengan demikian, dapat disimpulkan bahwa perubahan nilai pH dan amonia memiliki pengaruh yang lebih besar terhadap kualitas air dalam dataset.



Gambar 1. Distribusi Frekuensi Kondisi Data pada Setiap Parameter Sensor

Setelah semua data dari sensor telah diberi label, kualitas air dapat dibagi menjadi tiga kategori, yaitu Baik(0), Waspada(1), dan Buruk(2). Ini dilakukan agar memiliki target yang jelas untuk digunakan saat membuat model klasifikasi. Dengan adanya kategori ini, setiap data pengukuran dapat menunjukkan kondisi kualitas air akuarium berdasarkan karakteristik yang dimiliki. Distribusi hasil pelabelan selanjutnya dianalisis untuk mengetahui proporsi masing-masing kelas dalam dataset penelitian.



Gambar 2. Distribusi Hasil Pelabelan Kualitas Air Aquarium

Distribusi hasil pelabelan ditunjukkan pada Gambar 2. Dari total 710 data, sebanyak 363 data (51,13%) diklasifikasikan sebagai kualitas air Baik, 231 data (32,54%) masuk kategori Waspada, dan 116 data (16,34%) termasuk kategori Buruk. Distribusi ini menunjukkan bahwa sebagian besar kondisi air masih berada pada kategori baik, namun terdapat proporsi data kategori waspada dan buruk yang cukup signifikan sehingga dapat digunakan untuk melatih model klasifikasi secara efektif.

3.3. Pembagian Data

Sebelum pelatihan model dimulai, kumpulan data dipisahkan menjadi dua bagian, yaitu data untuk pelatihan dan data untuk pengujian dengan perbandingan 80:20 melalui teknik *train-test split*. Pemisahan data dilakukan secara *stratified* untuk memastikan bahwa proporsi masing-masing kelas tetap sama pada kedua bagian data tersebut.

Tabel 4. Contoh Dataset Parameter Air Aquarium

Kelas	Data Latih	Data Uji
Baik (0)	290	73
Waspada (1)	185	46
Buruk (2)	93	26
Total	568	142

Berdasarkan Tabel 4, sebanyak 568 data digunakan sebagai data pelatihan dan 142 data sebagai data pengujian. Penyebaran tiap kategori pada data pelatihan dan pengujian cukup stabil, sehingga data yang dipakai dianggap cukup mewakili untuk proses pelatihan dan evaluasi model LightGBM.

3.4. Hasil Hyperparameter Tuning Model LightGBM

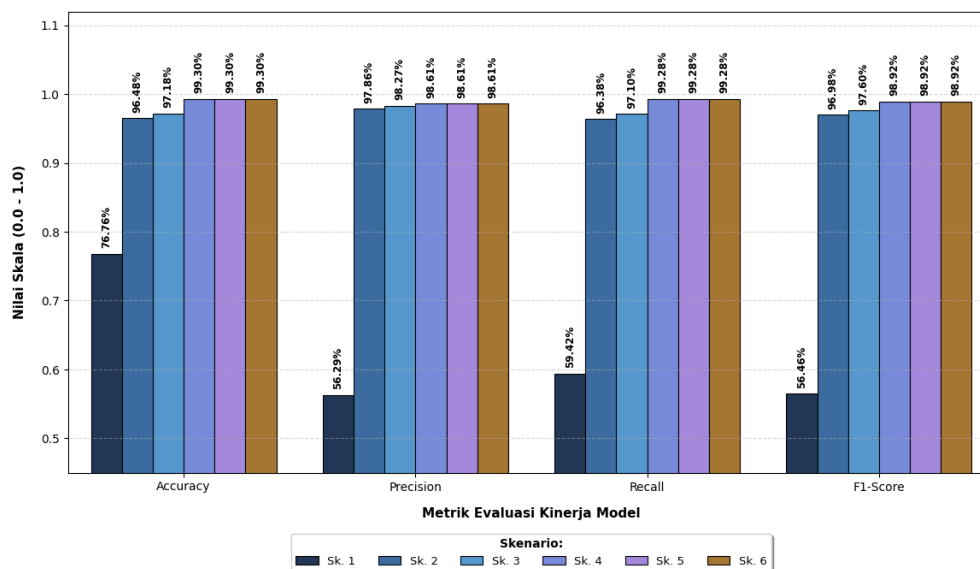
Dalam studi ini, dilakukan penyesuaian *hyperparameter* untuk memperoleh konfigurasi model LightGBM yang paling efektif dalam mengklasifikasikan kualitas air aquarium. Uji coba dilaksanakan dengan enam variasi parameter yang berbeda, termasuk nilai *learning rate*, *max depth*, *num leaves*, dan *n_estimators*. Perubahan parameter tersebut diterapkan untuk menilai dampak kompleksitas model terhadap kemampuan klasifikasi data. Setiap variasi kemudian dinilai menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* pada data pengujian untuk mengetahui pengaturan parameter yang memberikan hasil terbaik.

Berdasarkan Gambar 3, perubahan konfigurasi hyperparameter berdampak besar pada performa model LightGBM. Pada Skenario 1, hasilnya tidak terlalu baik dengan akurasi sebesar 76,76%, *precision* 56,29%, *recall* 59,42%, dan *F1-score* 56,46%. Ini berarti model LightGBM belum mampu memahami pola data dengan baik, sehingga masih banyak kesalahan dalam klasifikasi.

Peningkatan performa mulai terlihat pada Skenario 2 dan Skenario 3. Akurasi model meningkat menjadi 96,48% pada Skenario 2 dan 97,18% pada Skenario 3. Kenaikan ini juga diikuti oleh peningkatan nilai *precision*, *recall*, dan *F1-score*. Seluruh nilai tersebut berada di atas 96%. Hasil tersebut menunjukkan bahwa penyesuaian hyperparameter mampu meningkatkan kemampuan model. Model dapat lebih baik dalam mengenali karakteristik masing-masing kelas kualitas air. Dengan begitu, kesalahan prediksi menjadi lebih rendah.

Performa terbaik didapat pada Skenario 4. Di sini, model menunjukkan akurasi 99,30%, *precision* 98,61%, *recall* 99,28%, dan *F1-score* 98,92%. *Recall* yang mendekati 100% menunjukkan bahwa model mampu mengenali sebagian besar data pada kelas yang sesuai. *Precision* yang tinggi juga menunjukkan bahwa prediksi yang dihasilkan model memiliki tingkat kesalahan yang rendah. Hal ini menunjukkan bahwa

pengaturan *hyperparameter* pada Skenario 4 berhasil meningkatkan kemampuan model dalam mengenali pola kualitas air dan melakukan klasifikasi dengan baik. Konfigurasi *hyperparameter* pada Skenario 4 memberikan keseimbangan antara kemampuan model dalam mempelajari pola data dan kompleksitas model. Nilai *learning_rate* sebesar 0,05 membuat proses pembelajaran berlangsung secara bertahap sehingga pembaruan model menjadi lebih stabil. Parameter *max_depth* sebesar 5 dan *num_leaves* sebesar 10 mengatur struktur pohon keputusan agar model mampu menangkap hubungan antarp parameter kualitas air tanpa menghasilkan model yang terlalu kompleks. Selain itu, penggunaan *n_estimators* sebanyak 150 menentukan jumlah pohon yang dibangun dalam proses *boosting* untuk meningkatkan kemampuan model dalam membedakan kategori kualitas air.

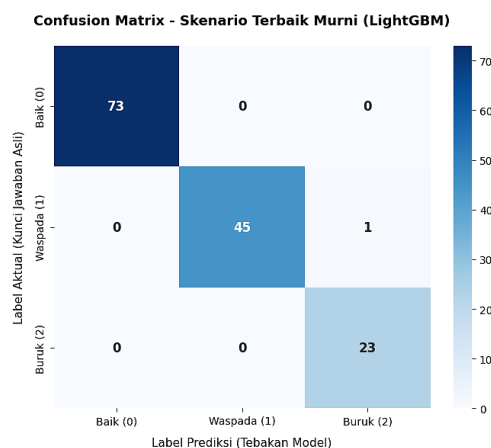


Gambar 3. Perbandingan Multi-Metrik Evaluasi Model LightGBM Berdasarkan 6 Skenario *Tuning*

Menariknya, Skenario 5 dan Skenario 6 menghasilkan hasil yang sama dengan Skenario 4. Meskipun Skenario 5 dan Skenario 6 menggunakan konfigurasi model yang lebih kompleks, tidak ada perubahan pada hasilnya. Ini berarti bahwa model sudah mencapai hasil terbaik pada Skenario 4, sehingga menambahkan kompleksitas model tidak membuat hasilnya lebih baik. Oleh karena itu, Skenario 4 dipilih sebagai konfigurasi terbaik dan digunakan pada tahap evaluasi selanjutnya.

3.5. Evaluasi Model Terbaik Menggunakan *Confusion Matrix*

Setelah diperoleh kombinasi *hyperparameter* terbaik pada Skenario 4, tahap berikutnya dilakukan pengujian performa model menggunakan *confusion matrix* guna melihat kemampuan algoritma dalam membedakan masing-masing kelas kualitas air. Metode ini digunakan dengan cara mengamati kesesuaian antara keluaran klasifikasi yang diberikan model terhadap kelas referensi pada data uji. Melalui proses tersebut, dapat diketahui jumlah hasil klasifikasi yang tepat serta pola kesalahan model dalam menentukan kategori kualitas air.



Gambar 4. *Confusion Matrix* Model LightGBM pada Skenario Terbaik

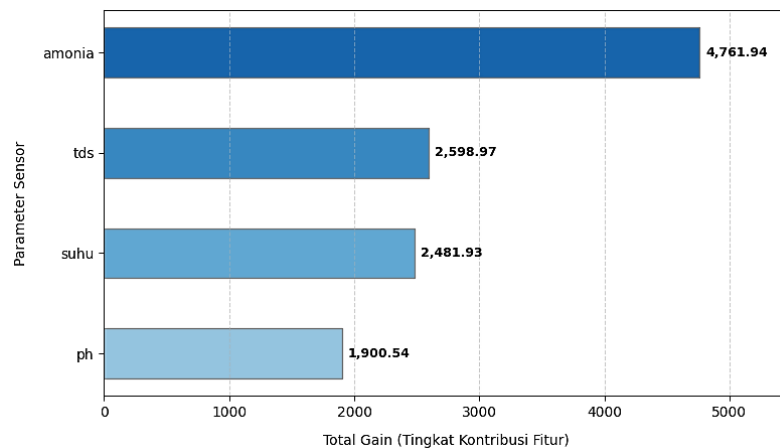
Setelah diperoleh konfigurasi *hyperparameter* terbaik pada Skenario 4, tahap evaluasi dilakukan dengan menerapkan *confusion matrix* untuk melihat tingkat ketepatan model dalam membagi data ke dalam setiap kategori kualitas air. Metode ini berfungsi sebagai alat pengukuran yang melihat kesesuaian antara keluaran klasifikasi dengan kelas referensi pada data pengujian, sehingga dapat diketahui hasil klasifikasi yang sesuai maupun pola kesalahan model pada tiap kategori kelas.

Berdasarkan Gambar 4, model LightGBM menunjukkan performa klasifikasi yang sangat baik pada data pengujian. Seluruh data pada kelas Baik (0) sebanyak 73 sampel berhasil diprediksi dengan benar tanpa kesalahan klasifikasi. Hal serupa juga terjadi pada kelas Buruk (2), di mana seluruh 23 sampel berhasil dikenali secara tepat oleh model. Pada kelas Waspada (1), sebanyak 45 dari 46 sampel berhasil diprediksi dengan benar, sedangkan 1 sampel diklasifikasikan sebagai kelas Buruk (2).

Secara keseluruhan, dari 142 data yang diuji, hanya ada satu kesalahan dalam prediksi. Sedikitnya jumlah kesalahan klasifikasi menunjukkan bahwa model ini mampu melakukan generalisasi dengan baik dalam mengategorikan kondisi kualitas air menjadi baik, waspada, dan buruk.

3.6. Analisis Kontribusi Parameter Sensor Menggunakan *Feature Importance*

Setelah mendapatkan model terbaik, perlu diketahui parameter sensor mana yang paling berpengaruh dalam menentukan kualitas air. Untuk itu, dilakukan analisis *feature importance*, analisis ini membantu memahami tingkat kontribusi masing-masing parameter sensor terhadap proses klasifikasi kualitas air. Pada algoritma LightGBM, tingkat kepentingan fitur dihitung menggunakan nilai *gain*, yaitu besarnya kontribusi suatu fitur dalam meningkatkan performa model saat proses pembentukan pohon keputusan. Semakin besar nilai *gain*, semakin besar pula pengaruh fitur tersebut terhadap hasil prediksi model.



Gambar 5. Hasil *Feature Importance* Model LightGBM

Berdasarkan Gambar 5, amonia adalah faktor yang paling berpengaruh dalam proses klasifikasi kualitas air, nilainya mencapai 4.761,94. TDS dan suhu juga penting dengan nilai masing-masing 2.598,97 dan 2.481,93. Sementara itu, pH memiliki kontribusi terendah, yaitu 1.900,54. Hasil ini menegaskan bahwa amonia sangat penting dalam menentukan kategori kualitas air pada dataset penelitian. Kadar amonia yang tinggi menunjukkan perubahan besar dalam kondisi kualitas air. Walaupun demikian, pH, suhu, dan TDS juga berperan penting karena semua parameter ini digunakan bersama dalam model LightGBM. Dengan demikian, amonia menjadi indikator utama dalam klasifikasi kualitas air akuarium pada penelitian ini.

4. KESIMPULAN

Hasil implementasi *Light Gradient Boosting Machine* (LightGBM) menunjukkan bahwa algoritma tersebut dapat diterapkan sebagai pendekatan klasifikasi dalam menentukan kondisi kualitas air akuarium berdasarkan parameter pH, suhu, amonia, dan TDS. Dataset penelitian ini mencakup sebanyak 710 sampel hasil pengamatan yang digunakan sebagai data analisis pengamatan. Sebelum digunakan, data tersebut telah melalui beberapa tahap, seperti pra-pemrosesan, pelabelan data, pembagian data, serta *hyperparameter tuning*. Model LightGBM ini digunakan untuk memahami pola hubungan antara parameter kualitas air dan mengelompokkan kondisi air akuarium menjadi tiga kategori, yaitu baik, waspada, dan buruk.

Berdasarkan pengujian beberapa kombinasi *hyperparameter* khusus, hasil terbaik diperoleh pada skenario 4 dengan *learning_rate* 0,05, *max_depth* 5, *num_leaves* 10, dan *n_estimators* 150. Hasil evaluasi menggunakan *confusion matrix* menunjukkan bahwa model ini dapat mencapai *accuracy* sebesar 99,30%, *precision* sebesar 98,61%, *recall* sebesar 99,28%, dan *F1-score* sebesar 98,92%. Hasil tersebut menunjukkan bahwa model LightGBM mampu memahami karakteristik data kualitas air dengan baik dan menghasilkan klasifikasi yang akurat dengan tingkat kesalahan prediksi yang rendah. Pencapaian tersebut menunjukkan

bahwa tujuan penelitian untuk membangun model klasifikasi kualitas air akuarium berbasis machine learning telah berhasil dilakukan, dengan model yang mampu menentukan kategori kondisi air berdasarkan parameter yang digunakan.

Penelitian ini menunjukkan bahwa algoritma LightGBM dapat digunakan untuk menentukan kondisi kualitas air akuarium secara otomatis berdasarkan data pengukuran. Model yang dibangun dapat membantu proses pemantauan kualitas air menjadi lebih cepat dan berbasis data, sehingga dapat memberikan informasi pendukung dalam pengelolaan kondisi lingkungan akuarium. Meskipun demikian, penelitian ini masih memiliki keterbatasan pada jumlah dataset dan cakupan kondisi lingkungan yang digunakan, sehingga pengujian lanjutan dengan jumlah data yang lebih besar dan variasi lingkungan akuarium yang lebih beragam masih diperlukan. Selain itu, penambahan parameter kualitas air lainnya dapat dipertimbangkan untuk meningkatkan kemampuan model dalam melakukan klasifikasi.

REFERENSI

- [1] K. Zhang *et al.*, “Water Quality Impact on Fish Behavior : A Review From an Aquaculture Perspective,” *Rev. Aquac.*, pp. 1–27, 2024, doi: 10.1111/raq.12985.
- [2] S. K. Nandi and A. Al Mamun, “Effects of Immediate Oxygen Supplementation (Sodium Carbonate and Hydrogen Peroxide) on Water Quality Parameters , Behavioural Responses and Survival of Puntius sophore Fingerlings,” *Vet. Med. Sci.*, vol. 11, no. 6, p. e70624, 2025, doi: 10.1002/vms3.70624.
- [3] M. M. Hemal *et al.*, “An Integrated Smart Pond Water Quality Monitoring and Fish Farming Recommendation Aquabot System,” *Sensors*, vol. 24, no. 11, p. 3682, 2024, doi: <https://doi.org/10.3390/s24113682>.
- [4] O. O. Olanubi, T. T. Akano, and O. S. Asaolu, “Design and development of an IoT - based intelligent water quality management system for aquaculture,” *J. Electr. Syst. Inf. Technol.*, vol. 11, 2024, doi: 10.1186/s43067-024-00139-z.
- [5] P. N. Bhowmik, K. Saini, N. Tagore, S. Priya, P. Anand, and B. Ate,s, “A Scalable Machine Learning Framework for Hydrological Water Quality Monitoring Using Physicochemical and Microbial Parameters,” *Water*, vol. 17, p. 2158, 2025, doi: <https://doi.org/10.3390/w17142158>.
- [6] G. Man, Q. Yun, Z. Qilin, L. Yuyong, and Z. Zhuoshi, “Surface water quality classification and prediction model based on multiple machine learning algorithms,” *Sci. Rep.*, vol. 15, p. 39907, 2025, doi: 10.1038/s41598-025-23775-5.
- [7] M. A. A. M. Hridoy *et al.*, “Advanced machine learning models for accurate water quality classification and WQI prediction : Implications for aquatic disease risk management,” *Sci. Total Environ.*, vol. 1008, p. 180965, 2025, doi: 10.1016/j.scitotenv.2025.180965.
- [8] T. Li, J. Lu, J. Wu, Z. Zhang, and L. Chen, “Predicting Aquaculture Water Quality Using Machine Learning Approaches,” *Water* 2022, vol. 14, no. 18, pp. 1–15, 2022, doi: <https://doi.org/10.3390/w14182836>.
- [9] M. M. Islam, M. A. Kashem, S. A. Alyami, and M. A. Moni, “Monitoring water quality metrics of ponds with IoT sensors and machine learning to predict fish species survival,” *Microprocess. Microsyst.*, vol. 102, p. 104930, 2023, doi: <https://doi.org/10.1016/j.micpro.2023.104930>.
- [10] P. Nuangpirom, S. Pitjarnit, V. Jaikampan, C. Peerakam, W. Nakkiew, and P. Jewpanya, “Machine Learning on Low-Cost Edge Devices for Real-Time Water Quality Prediction in Tilapia Aquaculture,” *Sensors*, vol. 25, no. 19, p. s25196159, 2025, doi: <https://doi.org/10.3390/s25196159>.
- [11] R. Baena-navarro, Y. Carriazo-regino, F. Torres-hoyos, and J. Pinedo-López, “Intelligent Prediction and Continuous Monitoring of Water Quality in Aquaculture : Integration of Machine Learning and Internet of Things for Sustainable Management,” *Water*, vol. 17, no. 82, pp. 1–25, 2025, doi: <https://doi.org/10.3390/w17010082>.
- [12] A. Nayan, J. Saha, A. N. Mozumder, K. R. Mahmud, A. K. Al Azad, and M. G. Kibria, “A Machine Learning Approach for Early Detection of Fish Diseases by Analyzing Water Quality,” *TRENDS Sci.*, vol. 18, no. 21, pp. 1–11, 2021, doi: <https://doi.org/10.48048/tis.2021.351> Focus to learn more.
- [13] G. Santhosh *et al.*, “An Efficient LightGBM-Based Machine Learning Framework for Water Quality Assessment in Urban Environments,” in *International Symposium on Sustainable Energy, Signal Processing and Cybersecurity (iSSSC)*, IEEE, 2025, pp. 0–5. doi: 10.1109/iSSSC66652.2025.11388640.
- [14] I. Budak, “Prediction of Water Quality’s pH value using Random Forest and LightGBM Algorithms,” *Memba Water Sci. J.*, vol. 11, no. 1, pp. 42–49, 2025, doi: <https://doi.org/10.58626/memba.1667338> Submission.
- [15] M. S. Bisht, “Water Quality Classification Using LightGBM: A Machine Learning Approach on a Large-Scale Dataset,” in *2025 IEEE International Conference on Recent Advances in Computing and Systems (REACS)*, India: IEEE, 2025, pp. 1–6. doi: 10.1109/REACS67479.2025.11413586.
- [16] P. Yasodha, “Data Preprocessing Methods for Machine Learning : An Empirical Comparison,” *Int. J. Multidiscip. Res.*, vol. 7, no. 3, pp. 1–7, 2025, [Online]. Available:

- <https://api.semanticscholar.org/CorpusID:279471746>
- [17] B. L. Ortiz *et al.*, “Data Preprocessing Techniques for AI and Machine Learning Readiness: Scoping Review of Wearable Sensor Data in Cancer Care,” *JMIR Mhealth Uhealth*, vol. 12, p. e59587, 2024, doi: 10.2196/59587.
 - [18] T. Gori, A. Sunyoto, and H. Al Fatta, “Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 1, pp. 215–224, 2024, doi: 10.25126/jtiik.20241118074.
 - [19] C. A. Indranie, A. Ramadhan, J. Raharjo, R. R. N. Ikhsan, M. F. Rohadi, and D. Ilham, “Impact of Train-Test Splitting Ratios on Machine Learning Models for Daily Electricity Load Time Series Forecasting,” in *2025 2nd International Conference on Information System and Information Technology (ICISIT)*, Yogyakarta: IEEE, 2025, pp. 1–6. doi: 10.1109/ICISIT66233.2025.11402981.
 - [20] M. Rosenblatt, L. Tejavibulya, R. Jiang, S. Noble, and D. Scheinost, “Data leakage inflates prediction performance in connectome-based machine learning models,” *Nat. Commun.*, vol. 15, pp. 1–15, 2024, doi: 10.1038/s41467-024-46150-w.
 - [21] H. Bichri, A. Chergui, and M. Hain, “Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 2, pp. 331–339, 2024, doi: <https://dx.doi.org/10.14569/IJACSA.2024.0150235>.
 - [22] N. F. Sahamony, Terttiaavini, and H. Rianto, “Analysis of Performance Comparison of Machine Learning Models for Predicting Stunting Risk in Children ’ s Growth,” *MALCOM Indones. J. Mach. Learn. Comput. Sci. J.*, vol. 4, no. 2, pp. 413–422, 2024, doi: <https://doi.org/10.57152/malcom.v4i2.1210>.
 - [23] E. Alotaibi and N. Nassif, “Artificial intelligence in environmental monitoring : in - depth analysis,” *Discov. Artif. Intell.*, vol. 4, no. 2, 2024, doi: 10.1007/s44163-024-00198-1.
 - [24] O. Jouini, K. Sethom, A. Namoun, N. Aljohani, M. H. Alanazi, and M. N. Alanazi, “A Survey of Machine Learning in Edge Computing : Techniques ,” *Technologies*, vol. 12, no. 6, pp. 1–34, 2024, doi: <https://doi.org/10.3390/technologies12060081>.
 - [25] G. Mohyuddin, M. A. Khan, A. Haseeb, S. Mahpara, M. Waseem, and A. M. Saleh, “Evaluation of Machine Learning approaches for precision Farming in Smart Agriculture System - A comprehensive Review,” *IEEE Access*, vol. 12, pp. 60155–60184, 2024, doi: 10.1109/ACCESS.2024.3390581.
 - [26] M. Milutinovic, “MACHINE LEARNING IN ENVIRONMENTAL MONITORING Mirjana Milutinovi ć,” *Facta Univ. Ser. Work. Living Environ. Prot.*, vol. 21, pp. 155–160, 2024, doi: <https://doi.org/10.22190/FUWLEP241029014M>.
 - [27] S. O. Ooko, L. Cheptegei, and S. M. Karume, “Application of machine learning for real-time water quality monitoring in developing countries : A review,” *Sustain. Futur.*, vol. 10, p. 100984, 2025, doi: 10.1016/j.sftr.2025.100984.
 - [28] A. Kriuchkova, V. Toloknova, and S. Drin, “Predictive model for a product without history using LightGBM: Pricing model for a new product,” *Mohyla Math. J.*, vol. 6, pp. 6–13, 2023, doi: 10.18523/2617-7080620236-13.
 - [29] S. Takahashi, A. Harada, and S. Yamada, “An Extended Closure Relation by LightGBM for Neutrino Radiation Transport in Core-collapse Supernovae,” *Astrophys. J.*, vol. 986, no. 1, 2025, doi: <https://doi.org/10.3847/1538-4357/add0ae>.
 - [30] C. Okolie, A. Adeleke, J. Smit, J. Mills, C. Ogbeta, and I. Maduako, “Performance analysis of Bayesian optimised gradient-boosted decision trees for digital elevation model (DEM) error correction : interim results,” *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. X-2–2024, no. June, pp. 179–183, 2024, doi: <https://doi.org/10.5194/isprs-annals-X-2-2024-179-2024>.
 - [31] L. Li, “LightGBM integration with modified data balancing and whale optimization algorithm for rock mass classification,” *Sci. Rep.*, vol. 14, no. 1, pp. 1–25, 2024, doi: 10.1038/s41598-024-73742-9.
 - [32] A. Ferdiansyah, F. R. Umbara, and F. Kasyidi, “Prediksi Penyakit Diabetes menggunakan Teknik Imputasi Missforest dan Klasifikasi LightGBM,” *MIND (Multimedia Artif. Intell. Netw. Database)*, vol. 10, no. 2, pp. 221–234, 2025, doi: <https://doi.org/10.26760/mindjournal.v10i2.221-234>.
 - [33] J. A. Ilemobayo *et al.*, “Hyperparameter Tuning in Machine Learning : A Comprehensive Review,” *J. Eng. Res. Reports*, vol. 26, no. 6, pp. 388–395, 2024, doi: 10.9734/jerr/2024/v26i61188.
 - [34] T. Kee and W. K. . Ho, “Optimizing Machine Learning Models for Urban Sciences : A Comparative Analysis of Hyperparameter Tuning Methods,” *Urban Sci.*, vol. 9, no. 9, p. 348, 2025, doi: 10.3390/urbansci9090348.
 - [35] Y. Rimal, N. Sharma, and A. Alsadoon, “The accuracy of machine learning models relies on hyperparameter tuning: - Student result classification using random forest, randomized search, grid search, bayesian, genetic, and optuna algorithms,” *Multimed. Tools Appl.*, vol. 83, no. 30, pp. 74349–74364, 2024, doi: 10.1007/s11042-024-18426-2.
 - [36] Z. Zuo, Y. Niu, J. Li, H. Fu, and M. Zhou, “Machine Learning for Advanced Emission Monitoring and Reduction Strategies in Fossil Fuel Power Plants,” *Appl. Sci.*, vol. 14, no. 18, p. 8442, 2024, doi:

- 10.3390/app14188442.
- [37] K.-T. Nguyen, T.-N. Tran, and H.-T. Nguyen, "Research on the Influence of Hyperparameters on the LightGBM Model in Load Forecasting," *Eng. Technol. Appl. Sci. Res*, vol. 14, no. 5, pp. 17005–17010, 2024, doi: <https://doi.org/10.48084/etasr.8266>.
- [38] F. Abbas *et al.*, "Optimizing Machine Learning Algorithms for Landslide Susceptibility Mapping along the Karakoram Highway , Gilgit Baltistan , Pakistan : A Comparative Study of Baseline , Bayesian , and Metaheuristic Hyperparameter Optimization Techniques," *Sensors (Basel)*., vol. 23, no. 15, p. 6843, 2023, [Online]. Available: <https://doi.org/10.3390/s23156843>
- [39] A. N. Hidayat, A. L. Ramdani, and L. Muthoharoh, "Aspect-Based Sentiment Analysis of Tourism Reviews Using Variants of the K-Nearest Neighbor Algorithm," *MALCOM Indones. J. Mach. Learn. Comput. Sci. J.*, vol. 6, no. 2, pp. 534–545, 2026, doi: doi.org/10.57152/malcom.v6i2.2522.