

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



Artikel Penelitian

Perbandingan Algoritma Random Forest dan XGBoost dalam Klasifikasi Kualitas Air Danau Maninjau

Muhammad Fajri ^{a,*}, Muhammad Rama Kencana ^b, Wendi Marsa Putra ^c, Muhammad Thoriq ^d

^{a,b,c,d} Program Studi Informatika, Universitas Adzka, Padang, 25175, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 28 Juni 2026

Revisi Akhir: 13 Agustus 2026

Diterbitkan Online: 25 Agustus 2026

KATA KUNCI

Danau Maninjau,
Kualitas Air,
Random Forest,
XGBoost,
Feature Alignment

KORESPONDENSI

E-mail: mfajri.tech@gmail.com*

ABSTRACT

Danau Maninjau merupakan salah satu danau prioritas nasional yang berada di Kabupaten Agam, Sumatera Barat. Selain unggul pada sektor pariwisata, Danau Maninjau dimanfaatkan oleh masyarakat sekitar untuk melakukan budidaya Keramba Jaring Apung (KJA). Namun lonjakan KJA yang tidak terkontrol menyebabkan tercemarnya perairan Danau Maninjau dengan status trofik eutrofik hingga hipereutrofik. Pemantauan kualitas perairan danau secara rutin dibutuhkan untuk menganalisis perkembangan kualitas perairan danau. Namun pemantauan berbasis laboratorium dinilai mahal dan membutuhkan waktu yang lama. Penelitian ini bertujuan membangun dan membandingkan model *machine learning* menggunakan algoritma *Random Forest* dan *XGBoost* dalam mengklasifikasi status mutu perairan Danau Maninjau. Penelitian ini memanfaatkan dataset indeks kualitas air (WQI) dengan skenario pembagian data 8:2. Hasil pengujian menunjukkan model *Random Forest* unggul secara signifikan dalam mengklasifikasi kualitas air dengan tingkat akurasi sebesar 97%, mengungguli model *XGBoost* yang hanya mencatatkan akurasi 87%. Implementasi model pada data pemantauan perairan Danau Maninjau berhasil mengidentifikasi secara konsisten status Kelas 3 (*Marginal*) pada empat lokasi krusial (Tengah Danau, Tanjung Sani, Sungai Tampang, dan PLTA Muko-Muko), yang diindikasikan kuat sebagai gejala eutrofikasi akibat supersaturasi oksigen terlarut (DO).

1. PENDAHULUAN

Danau Maninjau merupakan danau vulkanik yang terletak di Kabupaten Agam, Sumatera Barat. Danau ini termasuk satu dari lima belas danau prioritas nasional yang memiliki daya tarik wisata karena keindahan alamnya [1], [2]. Namun di balik keindahan tersebut, pemanfaatan ruang perairan danau selama beberapa dekade terakhir didominasi oleh aktivitas budidaya perikanan melalui Keramba Jaring Apung (KJA). Aktivitas ini pertama kali dilakukan pada tahun 1992 dan terus berkembang pesat hingga menjadi mata pencarian utama bagi masyarakat yang tinggal di selingkar Danau Maninjau [3].

Perkembangan KJA yang pesat ini sayangnya tidak diimbangi dengan pengendalian yang memadai. Menurut data yang dirilis

Bappeda Agam pada tahun 2022, jumlah KJA di danau Maninjau telah mencapai 23.359 petak [4]. Angka ini jauh melampaui daya tampung dan daya dukung danau yang telah ditetapkan sebesar 6.000 petak melalui Peraturan Daerah Nomor 5 Tahun 2014 [5]. Akibatnya, kondisi perairan Danau Maninjau mengalami pencemaran dengan status tercemar berat, serta status trofik eutrofik hingga hipereutrofik pada area tertentu [6], [7].

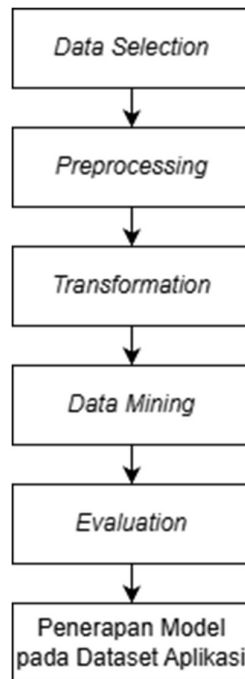
Dampak paling nyata dari kondisi ini adalah terjadinya kematian ikan massal yang terjadi hampir setiap tahun, yang oleh masyarakat setempat dikenal dengan istilah Tubo. Fenomena ini merujuk pada peristiwa *upwelling*, yaitu pergerakan massa air dari dasar danau menuju permukaan akibat perbedaan temperatur [8]. Pergerakan tersebut membawa zat-zat beracun yang berasal dari sisa metabolisme dan endapan pakan ikan di dasar danau, yang kemudian mencemari lapisan permukaan dan menyebabkan kematian puluhan hingga ribuan ton ikan pada Danau Maninjau

[9], [10]. Meskipun demikian, sebagian besar petani memilih untuk tetap bertahan dan melanjutkan usahanya, karena mereka menilai KJA masih menjadi sumber penghidupan yang paling realistis dibandingkan pilihan pekerjaan lainnya [8].

Meskipun dampak ekologi dari fenomena Tubo sudah masif, evaluasi terhadap kondisi perairan di danau Maninjau masih terbatas dan hanya dilaporkan pada waktu-waktu tertentu melalui edaran pemerintah maupun hasil penelitian. Proses penentuan kualitas air danau yang kompleks membutuhkan analisis mendalam yang terkadang bergantung pada pengujian laboratorium yang memakan waktu dan biaya [11], [12]. Hal ini menyebabkan kurangnya pelaporan kondisi perairan danau secara rutin, berkala, dan *real-time*, yang dapat diketahui secara langsung oleh masyarakat. Disamping itu, hasil data uji dari pemerintah daerah kabupaten Agam dan peneliti umumnya masih terbatas pada titik-titik sampling tertentu. Celah penelitian ini terletak pada tiadanya sistem interpretasi otomatis yang mampu menilai kondisi perairan Danau Maninjau berdasarkan parameter-parameter kimia hasil pengujian lapangan secara cepat, langsung, dan terstandarisasi.

Untuk menjembatani gap tersebut, penelitian ini menerapkan strategi *feature alignment* menggunakan dua sumber data yang diselaraskan secara parameter fisis dan kimiawi. Model kecerdasan buatan akan dilatih dan divalidasi terlebih dahulu menggunakan *Surface Water Quality Index Dataset* yang terdiri dari 7.790 baris data yang berasal dari Cork Harbour, Irlandia [13]. Penelitian ini menerapkan dua jenis algoritma *Machine Learning* dengan arsitektur berbeda untuk diuji performanya, yaitu *Random Forest Classifier* dan *XGBoost Classifier*. Setelah dievaluasi akurasi pada data global, model akan diimplementasikan sebagai instrumen pengujian pada data pemantauan kualitas air Danau Maninjau milik Pemkab Agam tahun 2018 [14]. Pendekatan ini diharapkan mampu memberikan penilaian otomatis terhadap status mutu air di berbagai titik sampel Danau Maninjau sekaligus membuktikan keandalan model prediktif global saat dihadapkan pada studi kasus ekologi lokal yang nyata.

2. METODE



Gambar 1. Diagram Alur Penelitian

Penelitian dilakukan untuk menganalisis kemampuan algoritma *Random Forest* dan *XGBoost* dalam mengklasifikasi kualitas perairan Danau Maninjau. Penelitian dilakukan secara bertahap menggunakan pendekatan *Knowledge Discovery in Databases* (KDD) yang terdiri atas *data selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation* [15]. Diagram alur penelitian ditampilkan pada Gambar 1.

2.1. Data Selection

Penelitian ini menggunakan dua data berbeda, yaitu data utama sebagai data latih dan evaluasi model, serta data aplikasi sebagai data penerapan model. Data utama diambil dari *website* Figshare yang merupakan data kualitas perairan pelabuhan Cork Harbour,

Irlandia, tahun 2007 – 2023, yang terdiri atas 7.790 baris dan 16 kolom [13]. Sedangkan data aplikasi merupakan data pemantauan kualitas perairan Danau Maninjau 2018 milik Pemerintah Daerah (Pemkab) Agam yang dirilis pada tahun 2022 [14].

Pada tahap ini, atribut pada kedua dataset diseleksi untuk memastikan atribut yang digunakan relevan dan tersedia pada kedua dataset. Hal ini dilakukan agar model yang dihasilkan dapat digunakan untuk memprediksi label pada dataset aplikasi.

2.2. Preprocessing

Preprocessing merupakan langkah awal mengolah data sebelum digunakan untuk membangun model. Langkah ini dilakukan

untuk membersihkan dan menghilangkan hal-hal yang tidak relevan pada data yang dapat menurunkan kualitas prediksi model [16].

Pada penelitian ini, *preprocessing* terdiri atas dua tahapan utama, yaitu penyelarasan fitur (*feature alignment*) antara dataset utama dan dataset aplikasi, serta melakukan *encoding* pada label atribut target.

Penyelarasan fitur dilakukan untuk memastikan kedua dataset memiliki fitur yang sama, sehingga model yang dihasilkan dari dataset utama dapat digunakan kembali pada dataset aplikasi. Sedangkan tahap *encoding* dilakukan untuk memastikan seluruh label pada data bersifat numerik agar mudah diproses saat melakukan pembangunan model.

2.3. Transformation

Proses *transformation* dilakukan untuk mengubah satuan label atribut DO pada dataset aplikasi agar sama dengan satuan label atribut DO pada dataset utama. Pada dataset aplikasi, atribut DO memiliki satuan konsentrasi massa (mg/L), sedangkan pada dataset utama memiliki satuan persentase kejenuhan (%). Untuk itu dilakukan transformasi satuan label DO pada dataset aplikasi dari satuan konsentrasi massa (mg/L) menuju persentase kejenuhan (%) menggunakan persamaan (1) [17].

$$DO_{\%} = \frac{DO_{mg/L}}{DO_{sat}} 100\% \quad (1)$$

dengan

$$DO_{sat} = 14.621 - 0.4477T + 0.0083T^2 - 0.000078T^3 \quad (2)$$

Transformasi juga dilakukan pada beberapa nilai di parameter BOD yang berada di bawah batas deteksi alat (<2.00 mg/L). untuk itu dilakukan penanganan data tersensor dengan menggunakan metode substitusi standar EPA dengan persamaan (3), sehingga nilai dikonversi menjadi 1.00 mg/L guna menjaga konsistensi input data pada model *machine learning*.

$$BOD_{new} = \frac{1}{2} LoD \quad (3)$$

2.4. Data Mining

Penelitian ini menggunakan dua algoritma berbeda untuk membangun model prediksi, yaitu algoritma *Random Forest* dan *XGBoost*. Kedua algoritma ini merupakan pengembangan dari algoritma *Decision Tree*.

Algoritma *Random Forest* bekerja dengan pohon keputusan paralel, di mana keputusan akhir dipengaruhi oleh banyaknya pohon yang memilih label tertentu. Algoritma ini memiliki mekanisme seleksi fitur bawaan yang mampu mengevaluasi *feature importance* dan memiliki atribut terbaik untuk meningkatkan keakuratan prediksi. Selain andal dalam menangani parameter data yang kompleks, pendekatan ini juga memberikan stabilitas kinerja yang tinggi meskipun model dihadapkan pada dataset yang mengandung *noise* atau ketidakseimbangan kelas [18], [19].

Berbeda dengan *Random Forest*, algoritma *XGBoost* bekerja dengan membangun pohon secara sekuensial, di mana pohon baru akan memperbaiki kelemahan dari pohon sebelumnya [20], [21]. Algoritma ini juga dilengkapi dengan regulasi untuk mencegah *overfitting*, sehingga sangat cocok diterapkan pada dataset besar [19].

2.5. Evaluation

Tahap evaluasi dilakukan untuk mengukur kemampuan model dalam memprediksi dan mengklasifikasikan data. Pada tahap ini, *Confusion Matrix* dimanfaatkan dengan beberapa parameter pengukuran, seperti *recall*, *accuracy*, serta *F1-score*. Metrik ini umum digunakan dalam mengukur kemampuan model dalam memprediksi suatu keputusan. Hasil evaluasi yang diperoleh kemudian digunakan dalam membandingkan kemampuan masing-masing model dalam memprediksi kualitas air.

2.6. Penerapan Model pada Data Aplikasi

Model yang dihasilkan selanjutnya diterapkan untuk memprediksi kualitas perairan Danau Maninjau berdasarkan data pemantauan kualitas air Danau Maninjau yang dilakukan oleh pemerintah kabupaten Agam pada tanggal 24 Agustus 2018 [14]. Tujuan dari penerapan ini ialah untuk mengetahui kondisi perairan Danau Maninjau, serta memberikan rancangan inovasi pemantauan bagi pemerintah kabupaten Agam agar selanjutnya pemantauan perairan danau dapat dilakukan secara berkala dengan memanfaatkan perkembangan teknologi.

3. HASIL

3.1. Karakteristik dan Distribusi Kelas Dataset Utama

Sebelum dilakukan proses pembangunan model, analisa terhadap karakteristik data pengembang perlu dilakukan guna memahami sebaran data. Tabel 1 menyajikan distribusi kelas target pada data utama.

Tabel 1. Distribusi Kelas Target pada Dataset Utama

Kelas	Jumlah
0	6
1	3864
2	184
3	3728
4	8
Total	7790

Diketahui bahwa dataset didominasi oleh dua kelas utama, yaitu kelas 1 (*Fair*) sebanyak 3.864 baris, dan 3 (*Marginal*) sebanyak 3.728 baris. Ketimpangan distribusi ini menjadi tantangan tersendiri dalam pembentukan batasan keputusan (*decision boundary*) bagi algoritma klasifikasi.

Pada dataset aplikasi, data yang diambil hanya data pemantauan kualitas perairan Danau Maninjau pada tanggal 24 Agustus 2018. Hal ini dilakukan karena data yang diambil pada waktu lain tidak memiliki atribut yang lengkap sesuai dengan kebutuhan atribut yang digunakan dalam penelitian ini.

3.2. Data Selection

Pada tahap ini dilakukan pemilihan atribut dari kedua dataset. Pemilihan atribut didasari atas ketersediaan data pada kedua

dataset. Atribut dibagi menjadi dua bagian, yaitu fitur dan target. Fitur yang digunakan dalam penelitian ini terdiri dari amonia, Kebutuhan Oksigen Biokimia (BOD), pH, suhu, dan oksigen terlarut (DO). Sedangkan target penelitian adalah klasifikasi kualitas air yang terdiri atas empat kelas, yaitu *Poor*, *Marginal*, *Fair*, *Good*, dan *Excellent*.

3.3. Preprocessing

Pada tahap *preprocessing*, dilakukan proses *feature alignment* dengan menyelaraskan atribut pada dataset utama dan dataset aplikasi. Penyelarasan ini dilakukan agar proses aplikasi model pada dataset aplikasi dapat dilakukan dengan baik. Tabel 2 menyajikan atribut yang digunakan pada kedua dataset.

Tabel 2. Atribut pada Kedua Dataset

Data Utama		Data Aplikasi	
Atribut	Satuan	Atribut	satuan
Ammonia-Total (as N)	mg/L	NH3 (mg/L)	mg/L
BOD - 5 days (Total)	mg/L	BOD (mg/L)	mg/L
pH		pH	
Temperature	°C	Temperatur (°C)	°C
Dissolved Oxygen	%	DO (mg/L)	mg/L
Label	Text	-	-

Selanjutnya dilakukan penyesuaian nama atribut pada setiap dataset yang disajikan pada Tabel 3, dan *encoding* pada label target *water_quality* yang disajikan pada Tabel 4.

Tabel 3. Penyesuaian Nama Atribut pada Kedua Dataset

Data Utama		Data Aplikasi	
Sebelum	Sesudah	Sebelum	Sesudah
Ammonia-Total (as N)	ammonia	NH3 (mg/L)	ammonia
BOD - 5 days (Total)	bod	BOD (mg/L)	bod
pH	ph	pH	ph
Temperature	temp	Temperatur (°C)	temp
Dissolved Oxygen	do	DO (mg/L)	do
Label	water_quality	-	-

Tabel 4. *Encoding* pada Label Target *water_quality*

Sebelum <i>Encoding</i>	Setelah <i>Encoding</i>
Excellent	0
Fair	1
Good	2
Marginal	3
Poor	4

3.4. Transformation

Pada tahap ini, dilakukan transformasi nilai pada atribut DO pada dataset aplikasi dari konsentrasi massa (mg/L) menjadi persentase kejenuhan (%), serta nilai BOD yang berada di bawah batas deteksi alat (< 2.00). Tabel 5 menampilkan perbandingan nilai DO dan BOD pada dataset aplikasi setelah dilakukan transformasi.

Tabel 5. Perbandingan Nilai DO dan BOD pada Dataset Aplikasi Sebelum dan Sesudah Transformasi

Sebelum		Sesudah	
DO (mg/L)	BOD (mg/L)	DO (%)	BOD (mg/L)
6.65	1.00	96.652144	1.00
6.55	1.00	95.198728	1.00
8.77	2.02	130.590331	2.02
9.27	2.42	134.731636	2.42
9.07	2.02	131.824804	2.02
7.06	2.02	100.162786	2.02

3.5. Pembangunan Model

Sebelum model dibangun, data utama dibagi menjadi dua bagian, yaitu data latih dan data uji. Pada tahap ini data dibagi menggunakan tiga rasio yang berbeda, yaitu 7:3, 8:2, dan 9:1. Rasio dengan hasil pengujian model terbaik dipilih untuk proses pembangunan model.

Tabel 6. Hasil Pengujian Tingkat Akurasi Model pada Masing-Masing Skenario Pembagian Data

Rasio Pembagian	Akurasi Random Forest	Akurasi XGBoost
7:3	0.948652	0.873770
8:2	0.969191	0.874198
9:1	0.984596	0.874198

Tabel 6 menyajikan tingkat akurasi masing-masing model berdasarkan rasio pembagian data. Hasil ini menunjukkan rasio 9:1 menghasilkan nilai akurasi tertinggi dibandingkan rasio lainnya. Namun pada penelitian ini, rasio 8:2 (80% data latih dan 20% data uji) diterapkan untuk menjaga konsistensi klasifikasi model. Data uji 10% pada skenario 9:1 berpotensi menampilkan performa yang tinggi secara semu, yang tidak menampilkan kemampuan generalisasi model yang sesungguhnya terhadap variasi data baru di lapangan. Selain itu, rasio 20% pada data uji dibutuhkan agar setiap kelas minoritas tetap mendapatkan alokasi jumlah sampel yang cukup dan seimbang.

3.6. Data Mining

Sebelum memasuki tahap pelatihan model, kedua algoritma dikonfigurasi menggunakan parameter dasar tertentu guna memastikan proses komparasi berjalan secara objektif dan setara (*fair comparison*). Rincian parameter yang digunakan untuk membangun model *Random Forest* dan *XGBoost* disajikan pada Tabel 7.

Tabel 7. Konfigurasi Parameter Model *Random Forest* dan *XGBoost*

Algoritma	Parameter	Nilai
Random Forest	n_estimator	100
	random_state	42
XGBoost	n_estimator	100
	learning_rate	0.1
	max_depth	6
	random_state	42
	eval_metric	mlogloss

3.7. Evaluation

3.7.1. Visualisasi Confusion Matrix Model

Visualisasi *confusion matrix* digunakan untuk menganalisis kemampuan model dalam mengklasifikasikan data uji. Perbedaan performa yang signifikan dapat dilihat pada visualisasi *confusion matrix* kedua model.

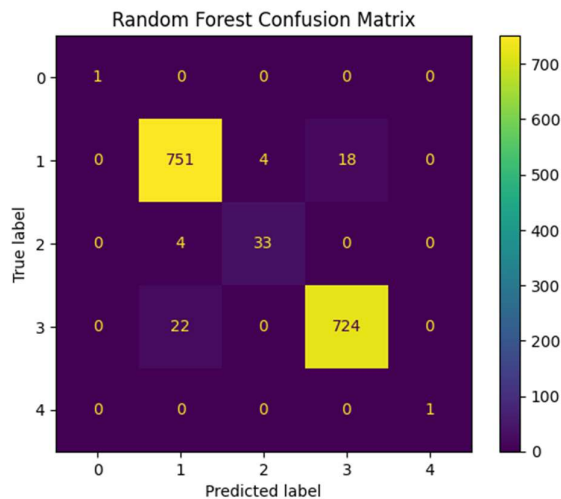
Gambar 2 menunjukkan visualisasi *confusion matrix* yang diperoleh oleh model *Random Forest*. Pada diagonal utama, model ini berhasil memprediksi dengan benar 1.510 dari 1.558 data uji. Hal yang berbeda ditampilkan pada gambar 3, di mana model *XGBoost* hanya mampu memprediksi 1.362 data dengan benar.

3.7.2. Feature Importance

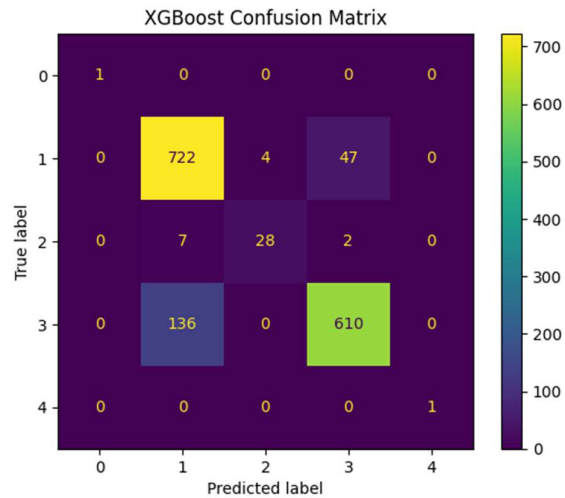
Berdasarkan hasil evaluasi tingkat kepentingan fitur (*feature importance*) pada kedua model, ditemukan bahwa fitur *Dissolved Oxygen (DO)* menjadi fitur yang memberikan pengaruh signifikan terhadap keputusan klasifikasi model pada angka 0.48 pada *Random Forest* dan 0.54 pada *XGBoost*. Angka ini jauh lebih tinggi dibandingkan fitur lainnya yang berada di angka 0.87 – 0.15 pada kedua model.

Tabel 8. *Feature Importance* pada Kedua Model

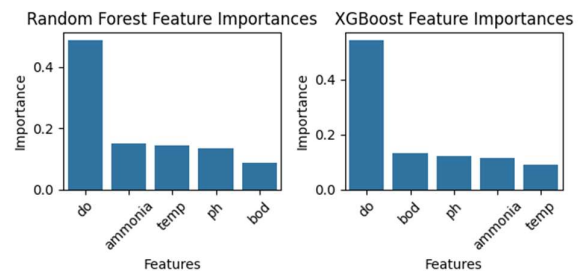
Fitur	Random Forest		XGBoost	
	nilai	Urutan	nilai	Urutan
do	0.541802	1	0.486543	1
bod	0.130993	2	0.087188	5
ph	0.121884	3	0.133608	4
ammonia	0.115680	4	0.149188	2
temp	0.089642	5	0.143473	3



Gambar 2. Visualisasi *Confusion Matrix* pada Model Random Forest



Gambar 3. Visualisasi *Confusion Matrix* pada Model XGBoost



Gambar 4. Visualisasi *Feature Importance* pada Kedua Model

4. PEMBAHASAN

4.1. Analisis Perbandingan Performa Model

Untuk mengetahui model terbaik dalam mengklasifikasikan kualitas air, dilakukan pengujian komparatif terhadap dua model yang telah dibangun dengan skenario data *split* 8:2. Hasil evaluasi menyeluruh dari kedua model tersebut disajikan pada Tabel 9.

Tabel 9. Perbandingan Performa Model *Random Forest* dan *XGBoost*

Model	Metrik	Weighted Avg	Macro Avg
Random Forest	Accuracy	0.97	-
	Precision	0.97	0.97
	Recall	0.97	0.97
	F1-Score	0.97	0.97
XGBoost	Accuracy	0.87	-
	Precision	0.88	0.93
	Recall	0.87	0.90
	F1-Score	0.87	0.91

Berdasarkan data yang disajikan pada Tabel 9, diketahui bahwa model *Random Forest* menunjukkan performa yang superior dengan perolehan akurasi sebesar 97%. Angka ini unggul 10% dibandingkan *XGBoost* yang hanya memperoleh akurasi sebesar 87%.

Indikator paling krusial dapat dilihat pada nilai *Macro Average* dan *Weighted Average* milik *Random Forest* yang konsisten di angka 0.97. hal ini mengindikasikan kekuatan model dalam

menggeneralisir seluruh kelas secara seimbang tanpa mengalami bias, meskipun dataset yang digunakan memiliki tingkat ketimpangan kelas (*imbalanced data*) yang cukup tinggi.

Kelemahan model *XGBoost* dapat dilihat pada tingginya angka salah tebak (*misclassification*) pada dua kelas mayoritas, yaitu kelas 1 (*Fair*) dan kelas 3 (*Marginal*), yang ditampilkan pada

Gambar 3. Pada model *XGBoost*, terdapat 136 data yang merupakan kelas 3 namun diklasifikasikan sebagai kelas 1, serta 47 data kelas 1 yang diklasifikasikan sebagai kelas 3. Sementara itu model *Random Forest* menampilkan hasil klasifikasi yang jauh lebih baik, dengan total 40 data meleset antara kelas 1 dan kelas 3.

Tabel 10. Hasil Evaluasi Per-Kelas Model *Random Forest* dan *XGBoost*

Kelas	RF Precision	RF Recall	RF F1-Score	XGB Precision	XGB Recall	XGB F1-Score	Support
0	1.00	1.00	1.00	1.00	1.00	1.00	1
1	0.97	0.97	0.97	0.83	0.93	0.88	773
2	0.89	0.89	0.89	0.88	0.76	0.81	37
3	0.98	0.97	0.97	0.93	0.82	0.87	746
4	1.00	1.00	1.00	1.00	1.00	1.00	1

4.2. Performa Model terhadap Kelas Minoritas

Pada pengujian terhadap kelas minoritas, khususnya kelas 2 (*Good*) dengan total 37 sampel data uji, model *Random Forest* kembali menunjukkan keunggulannya. Model *Random Forest* berhasil memperoleh nilai *F1-Score* sebesar 0.89 pada kelas 2, sementara model *XGBoost* hanya memperoleh nilai *F1-Score* sebesar 0.81, yang dipicu oleh rendahnya nilai *recall* sebesar 0.76. Hal ini mengindikasikan bahwa model *XGBoost* gagal mendeteksi 24% data kualitas air kelas 2 (*Good*), dan justru diklasifikasikan pada kelas lain.

Adapun untuk kelas 0 (*Excellent*) dan kelas 4 (*Poor*), kedua model mendapatkan nilai sempurna 1.00. Namun perolehan ini tidak dapat dijadikan bahan diskusi lebih dalam karena jumlah sampel pengujian masing-masing kelas sangat sedikit (1 sampel per-kelas). Secara umum, pendekatan *Bagging* (*Random Forest*) yang merata-ratakan keputusan dari ratusan pohon acak terbukti

jauh lebih stabil dan efektif untuk dataset kualitas air ini dibandingkan pendekatan *Boosting* (*XGBoost*).

4.3. Analisis Signifikansi Fitur

Berdasarkan hasil evaluasi tingkat kepentingan fitur (*feature importance*) pada kedua model (Tabel 8), ditemukan bahwa fitur *Dissolved Oxygen* (DO) menjadi fitur yang memberikan pengaruh signifikan terhadap keputusan klasifikasi model. Dominasi fitur ini sejalan dengan temuan berbasis *SHAP value* yang dilakukan oleh Arzhanghi dan Partani (2026) yang menunjukkan bahwa DO sebagai fitur paling berpengaruh berdasarkan analisis yang dilakukan [22].

Keberhasilan kedua model dalam menempatkan DO sebagai fitur utama membuktikan bahwa model komputasi yang dibangun tidak berjalan secara acak (*black box*). Sebaliknya, model telah berhasil mengadopsi cara berpikir pakar lingkungan yang menempatkan DO sebagai parameter tertinggi, sebelum kemudian mengevaluasi parameter pendukung lainnya seperti temperatur dan BOD [22].

Tabel 11. Hasil prediksi kualitas perairan Danau Maninjau Menggunakan *Random Forest* dan *XGBoost*

Lokasi	Temperatur	PH	DO (ml/L)	DO (%)	BOD	Ammonia	Prediksi Random Forest	Prediksi XGBoost
Bayur	28	5.7	6.65	96.65	1.00	0.6	1 (Fair)	3 (Marginal)
Sungai batang	28	6.5	6.55	95.20	1.00	0.8	1 (Fair)	3 (Marginal)
Tengah danau	29	8.0	8.77	130.59	2.02	0.8	3 (Marginal)	3 (Marginal)
Tanjung sani	28	7.7	9.27	134.73	2.42	0.9	3 (Marginal)	3 (Marginal)
Sungai tampang	28	7.1	9.07	131.82	2.02	0.8	3 (Marginal)	3 (Marginal)
PLTA Muko-muko	27	7.2	7.06	100.16	2.02	1.3	3 (Marginal)	3 (Marginal)

4.4. Implementasi dan Analisis Prediksi Kualitas Air Danau Maninjau

Setelah kedua model dievaluasi pada tahap pengujian, selanjutnya dilakukan implementasi kedua model untuk mengklasifikasi kualitas perairan Danau Maninjau berdasarkan kondisi riil di lapangan. Proses klasifikasi ini dilakukan dengan menggunakan data milik Pemerintah Daerah (Pemda) Agam yang mencakup enam titik pengamatan strategis, yaitu Bayur, Sungai Batang, Tengah Danau, Tanjung Sani, Sungai Tampang, dan PLTA Muko-Muko. Hasil pemetaan kualitas air Danau Maninjau dari kedua model disajikan pada Tabel 11.

Berdasarkan hasil klasifikasi yang disajikan pada Tabel 11, diketahui bahwa kedua model memiliki konsistensi tinggi dalam menetapkan status Kelas 3 (*Marginal*) pada empat lokasi krusial, yaitu Tengah Danau, Tanjung Sani, Sungai Tampang, dan PLTA Muko-Muko.

Secara hidrologis, keputusan mutlak kedua model dilandasi oleh deteksi nilai persentase DO yang sangat tinggi di angka 130.59% sampai 134.73%, bersamaan dengan kenaikan parameter beban organik BOD (2.02 hingga 2.42). Integrasi model komputasi dengan teori dari Zhao dan Chen (2025) mengonfirmasi bahwa kondisi ini merupakan indikator kuat bahwa Danau Maninjau sedang mengalami gejala eutrofikasi [23]. Lonjakan kadar oksigen terlarut yang melampaui batas jenuh (supersaturasi) dipicu oleh aktivitas fotosintesis masif dari ledakan populasi alga

(*blooming algae*). Kondisi ini menimbulkan dampak negatif terhadap danau. Di mana saat alga mati, proses dekomposisi bangkai alga oleh bakteri pengurai akan mengonsumsi oksigen yang ada di perairan secara masif, sehingga terjadi penurunan konsentrasi oksigen dalam air. Selain itu permukaan air yang tertutupi alga akan menghalangi masuknya sinar matahari yang mengakibatkan terganggunya proses fotosintesis di dasar danau [6].

Perbedaan hasil klasifikasi kedua model terdeteksi pada wilayah Bayur dan Sungai Batang, di mana *Random Forest* menetapkan status Kelas 1 (*Fair*) sementara *XGBoost* menetapkan Kelas 3 (*Marginal*). Perbedaan ini menunjukkan perbedaan karakteristik model dalam menentukan pola klasifikasi.

Pada kedua lokasi tersebut, kondisi air dicirikan oleh nilai pH yang cenderung mengarah pada kondisi asam (5.7 dan 6.5) serta temperatur yang cukup hangat sebesar 28°C. Model *XGBoost* mengaitkan pH asam dan suhu hangat sebagai ancaman penurunan kualitas air, meskipun kadar DO berada dalam batas normal. Pendekatan berbeda dilakukan oleh model *Random Forest*, di mana melalui mekanisme *ensemble voting*, fitur utama berupa DO dan beban organik (BOD) yang berada dalam batas aman mendasari keputusan kualitas air danau berada dalam kategori Kelas 1 (*Fair*). Mengingat nilai akurasi dan *F1-Score* model *Random Forest* lebih tinggi pada tahap pengujian, maka keputusan status *Fair* pada wilayah Bayur dan Sungai Batang dinilai lebih objektif dan sesuai dengan kondisi riil di lapangan.

5. KESIMPULAN

Berdasarkan hasil pengujian komparatif antara model *Random Forest* dan *XGBoost* dalam mengklasifikasikan kualitas air Danau Maninjau menggunakan *feature alignment*, dapat ditarik beberapa kesimpulan bahwa algoritma *Random Forest* terbukti menjadi model terbaik dan kokoh dalam mengklasifikasikan kualitas air berdasarkan data utama. Dengan skenario pembagian data latih dan uji 8:2, *Random Forest* berhasil mendapatkan nilai akurasi sebesar 97%, jauh mengungguli *XGBoost* yang hanya meraih akurasi 87%.

Algoritma *Random Forest* menunjukkan keandalan dalam menangani ketimpangan kelas pada dataset. Hal ini dibuktikan dengan perolehan nilai rata-rata makro yang konsisten di angka 0.97, serta kemampuannya dalam mengenal kelas minoritas seperti Kelas 2 (*Good*) dengan *F1-Score* sebesar 0.89. Perolehan ini jauh lebih tinggi dibandingkan algoritma *XGBoost* yang hanya memperoleh *F1-Score* di angka 0.81 akibat rendahnya nilai *recall*.

Parameter *Dissolved Oxygen* (DO) menjadi fitur yang memiliki pengaruh terbesar pada keputusan pada kedua model klasifikasi. Implementasi model pada data pemantauan kualitas perairan milik Pemerintah Daerah (Pemda) Agam menunjukkan kondisi perairan yang cukup buruk (*Marginal*) pada daerah Tengah Danau, Tanjung Sani, Sungai Tampang, dan PLTA Muko-Muko. Hal ini didasari dengan lonjakan DO ekstrem yang mencapai 130.59% – 134.73% yang diikuti dengan kenaikan beban pencemaran organik (BOD). Sementara itu untuk wilayah Bayur dan Sungai Batang, keputusan *Random Forest* yang menerapkan

status Kelas 1 (*Fair*) dinilai lebih objektif dan sesuai kondisi riil dibandingkan *XGBoost*.

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan dalam interpretasi hasil yaitu distribusi kelas pada dataset utama yang digunakan untuk melatih dan menguji model bersifat tidak seimbang, di mana Kelas 0 (*Excellent*) dan Kelas 4 (*Poor*) hanya diwakili oleh masing-masing satu sampel pada data uji, sehingga nilai metrik pada kedua kelas tersebut tidak dapat dijadikan acuan yang representatif.

Disamping itu, model latih menggunakan data kualitas perairan dari Cork Harbour, Irlandia, yang mungkin memiliki karakteristik ekosistem berbeda dengan perairan tropis seperti Danau Maninjau. Meskipun pendekatan *feature alignment* telah diterapkan untuk menjembatani perbedaan ini, validasi lintas domain yang lebih komprehensif tetap diperlukan untuk memastikan keandalan model pada konteks ekologi lokal. Data aplikasi pemantauan perairan Danau Maninjau yang digunakan hanya terbatas pada satu tanggal pengamatan, yaitu 24 Agustus 2018, karena keterbatasan atribut pada tanggal pengamatan lainnya.

Berdasarkan kesimpulan dan keterbatasan dari penelitian ini, disarankan bagi penelitian selanjutnya untuk memperluas data lapangan guna menjaga distribusi kelas pada data latih model. Selain itu ketersediaan dataset lokal untuk melatih model lebih diprioritaskan untuk menjamin keselarasan karakteristik perairan dengan Danau Maninjau.

DAFTAR PUSTAKA

- [1] *Peraturan Presiden (Perpres) Nomor 60 Tahun 2021 tentang Penyelamatan Danau Prioritas Nasional*, vol. No. 60 Tahun 2021. 2021.
- [2] S. Purwanto, S. Yudistira, dan L. Asnur, "Penguatan Kapasitas Sumber Daya Manusia dalam Pengelolaan Atraksi Wisata Air di Danau Maninjau," vol. 6, no. 1, 2026.
- [3] Junaidi dkk., "Floating cage aquaculture production in Indonesia: Assessment of opportunities and challenges in Lake Maninjau," *AIMSES*, vol. 9, no. 1, hlm. 1–15, 2022, doi: [10.3934/environsci.2022001](https://doi.org/10.3934/environsci.2022001).
- [4] A. Ridoh, "Website Resmi Badan Perencanaan Pembangunan Daerah Kabupaten Agam | Data Keramba Jaring Apung di Danau Maninjau Tahun 2022." Diakses: 10 Juni 2026. [Daring]. Tersedia pada: <https://bappeda.agamkab.go.id/View-dokumen/data-keramba-jaring-apung-di-danau-maninjau-tahun-2022.html>
- [5] *Peraturan Daerah (Perda) Kabupaten Agam Nomor 5 Tahun 2014 tentang Pengelolaan Kelestarian Kawasan Danau Maninjau*, vol. No. 5 Tahun 2014. 2014.
- [6] A. Sunaryani, "Penentuan Status Mutu Air dan Status Trofik di Perairan Danau Maninjau: Determination of Water Quality Status and Trophic Classification of Lake Maninjau," *Jurtekling*, vol. 24, no. 1, hlm. 021–027, Jan 2023, doi: [10.55981/jtl.2023.246](https://doi.org/10.55981/jtl.2023.246).
- [7] H. Syandri, A. Azrita, dan A. Mardiah, "Water Quality Status and Pollution Waste Load from Floating Net Cages at Maninjau Lake, West Sumatera Indonesia," *IOP Conf.* <https://doi.org/10.25077/TEKNOSI.v12i2.2026.276-283>

- Ser.: Earth Environ. Sci.*, vol. 430, no. 1, hlm. 012031, Jan 2020, doi: [10.1088/1755-1315/430/1/012031](https://doi.org/10.1088/1755-1315/430/1/012031).
- [8] S. Silvia dan R. Resdati, "Dampak Tubo Belerang terhadap Sosial Ekonomi Masyarakat Kawasan Danau Maninjau Nagari Koto Malintang Kecamatan Tanjung Raya Kabupaten Agam," *Etnorefika*, vol. 14, no. 3, hlm. 486–505, Okt 2025, doi: [10.33772/etnorefika.v14i3.2878](https://doi.org/10.33772/etnorefika.v14i3.2878).
- [9] A. N. Agency, "Ikan keramba jaring apung Danau Maninjau mati lima ton selama 2026," *Antara News Sumbar*. Diakses: 10 Juni 2026. [Daring]. Tersedia pada: <https://sumbar.antaranews.com/berita/765535/ikan-keramba-jaring-apung-danau-maninjau-mati-lima-ton-selama-2026>
- [10] S. Devi dan R. L. Arios, "Tubo Dalam Perspektif Ekologi Budaya Petani Keramba Jaring Apung Di Kawasan Danau Maninjau Propinsi Sumatera Barat," *PATANJALA*, vol. 13, no. 1, hlm. 35, Apr 2021, doi: [10.30959/patanjala.v13i1.646](https://doi.org/10.30959/patanjala.v13i1.646).
- [11] C. Henny dan S. Nomosatryo, "Changes in water quality and trophic status associated with cage aquaculture in Lake Maninjau, Indonesia," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 31, hlm. 012027, Jan 2016, doi: [10.1088/1755-1315/31/1/012027](https://doi.org/10.1088/1755-1315/31/1/012027).
- [12] J. Xu *dkk.*, "Assessing and predicting water quality index with key water parameters by machine learning models in coastal cities, China," *Heliyon*, vol. 10, no. 13, hlm. e33695, Jun 2024, doi: [10.1016/j.heliyon.2024.e33695](https://doi.org/10.1016/j.heliyon.2024.e33695).
- [13] A. Rahman *dkk.*, "Surface Water Quality Index (WQI) Forecasting Dataset." figshare, hlm. 714018 Bytes, 2025. doi: [10.6084/M9.FIGSHARE.28184252.V1](https://doi.org/10.6084/M9.FIGSHARE.28184252.V1).
- [14] A. Ridoh, "Website Resmi Badan Perencanaan Pembangunan Daerah Kabupaten Agam | Hasil Uji Kualitas Air Danau Maninjau Kabupaten Agam Tahun 2018 s/d 2020." Diakses: 19 Juni 2026. [Daring]. Tersedia pada: <https://bappeda.agamkab.go.id/View-dokumen/hasil-uji-kualitas-air-danau-maninjau-kabupaten-agam-tahun-2018-s-d-2020.html>
- [15] M. Thoriq, F. Maulana, Y. S. Eirlangga, N. Hayati, dan M. A. Madani, "Implementasi Algoritma Naïve Bayes dalam Prediksi Penerimaan Mahasiswa Penerima Beasiswa KIP di Universitas Adzkia," *JURNAL FASILKOM*, vol. 15, no. 1, hlm. 108–114, Apr 2025, doi: [10.37859/jf.v15i1.8959](https://doi.org/10.37859/jf.v15i1.8959).
- [16] R. Oktafiani dan R. Rianto, "Perbandingan Algoritma Support Vector Machine (SVM) dan Decision Tree untuk Sistem Rekomendasi Tempat Wisata," *TEKNOSI*, vol. 9, no. 2, hlm. 113–121, Agu 2023, doi: [10.25077/TEKNOSI.v9i2.2023.113-121](https://doi.org/10.25077/TEKNOSI.v9i2.2023.113-121).
- [17] A. Fauzan, W. R. Melani, dan T. Apriadi, "Tingkat Kesuburan Perairan di Perairan Tembeling Tanjung, Kabupaten Bintan, Provinsi Kepulauan Riau," *Manajemen Sumberdaya Perairan*, vol. 2, no. 1, hlm. 22–28, Nov 2018, doi: [10.31629/akuatiklestari.v2i1.2349](https://doi.org/10.31629/akuatiklestari.v2i1.2349).
- [18] A. Firmansyah, M. F. Syahidin, dan Y. S. Triana, "Prediksi Kebakaran Hutan Berdasarkan Titik Panas dan Iklim Menggunakan Algoritma Random Forest," *TEKNOSI*, vol. 10, no. 2, hlm. 145–155, Sep 2024, doi: [10.25077/TEKNOSI.v10i2.2024.145-155](https://doi.org/10.25077/TEKNOSI.v10i2.2024.145-155).
- [19] M. Firmansyah, A. A. Rudiyyat, R. Purnama, H. P. Utomo, dan H. S. Firmansyah, "Penerapan Algoritma Xgboost dan Random Forest dalam Prediksi Uraian Resiko Proyek Pada Dinas XYZ," vol. 20, no. 2, Mar 2026.
- [20] E. Mustika Sari, C. Sabila, R. Fakhrizal Adam, dan R. Kurniawan, "Analisis dan Prediksi Indeks Kualitas Udara Jakarta: Penerapan Algoritma XGBoost," *TEKNOSI*, vol. 11, no. 2, hlm. 161–169, Sep 2025, doi: [10.25077/TEKNOSI.v11i2.2025.161-169](https://doi.org/10.25077/TEKNOSI.v11i2.2025.161-169).
- [21] F. I. Sari, E. L. Gunawan, dan C. A. Adhigiadany, "Model Prediksi Kepadatan Lalu Lintas: Perbandingan Antara Algoritma Random Forest dan XGBoost," 2023.
- [22] A. Arzhangi dan S. Partani, "Water quality index prediction via a robust machine learning model using oxygen-related indices for river water quality monitoring," *Sci Rep*, vol. 16, no. 1, hlm. 6102, Jan 2026, doi: [10.1038/s41598-026-36156-3](https://doi.org/10.1038/s41598-026-36156-3).
- [23] Y. Zhao dan M. Chen, "Prediction of river dissolved oxygen (DO) based on multi-source data and various machine learning coupling models," *PLOS ONE*, vol. 20, no. 3, hlm. e0319256, Mar 2025, doi: [10.1371/journal.pone.0319256](https://doi.org/10.1371/journal.pone.0319256).

NOMENKLATUR

$DO_{\%}$	Persentase Kejenuhan Oksigen Terlarut
$DO_{mg/L}$	Konsentrasi Oksigen Terlarut
DO_{sat}	Tingkat Oksigen Terlarut Jenuh
BOD_{new}	Kebutuhan Oksigen Biokimia Baru
LoD	Batas Deteksi Alat