

**Aplikasi *Machine Learning* Untuk Prediksi Status Perairan Sungai Ogan
Berbasis Struktur Komunitas Ikan**

**Machine Learning Application for Prediction of Ogan River Waters Status
Based on Fish Community Structure**

**Allamanda Catharica^{1*}, Muhammad Nizar², Ria Fahleny¹, Nadila Nur Khotimah¹,
Baref Agung Wicaksono¹**

¹Program Studi Manajemen Sumberdaya Perairan, Fakultas Perikanan
Universitas Islam Ogan Komering Ilir, Kayuagung

Jl. Sulaiman Raden Anom No.333 Celikah Muara Baru Kayuagung, Ogan Komering Ilir

²Program Studi Akuakultur, Fakultas Pertanian, Universitas Muhammadiyah Palembang,
Jl. Jenderal Ahmad Yani 13, 1 Ulu, Seberang Ulu II, Palembang, Sumatera Selatan 30129

*Korespondensi email: catharica.msdp.uniski@gmail.com

ABSTRACT

This study aims was to develop a machine learning prediction model to evaluate the status of water in the Ogan River, in order to inform conservation efforts and effective policy decisions. The model was developed using Random Forest, XGBoost, and Logistic Regression algorithms, with inputs consisting of geographical features (longitude and latitude) and diversity indices (H' , J' , D). The modeling results show that all three algorithms achieved perfect classification performance (accuracy of 1.00) on the test data. Further interpretative analysis indicates that the water quality along the Ogan River in Sungai Pinang District has entered an unhealthy condition, particularly in the river mouth area. Random Forest was identified as the best model due to its feature importance interpretation, supported by SHAP analysis, which is most relevant to ecological understanding. These findings can serve as a basis for more effective water resource management.

Keywords: *Diversity Index, Machine Learning, Random Forest, Ogan River, Water Status*

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan model prediksi *machine learning* untuk mengevaluasi status perairan di Sungai Ogan, guna menginformasikan upaya konservasi dan pengambilan keputusan kebijakan yang efektif. Model dikembangkan menggunakan algoritma *Random Forest*, *XGBoost*, dan *Logistic Regression*, dengan input berupa fitur geografis (*longitude* dan *latitude*) dan indeks diversitas (H' , J' , D). Hasil permodelan menunjukkan bahwa ketiga algoritma mencapai performa klasifikasi yang sempurna (akurasi 1.00) pada data uji. Analisis interpretatif lebih lanjut menunjukkan bahwa status perairan di sepanjang Sungai Ogan di Kecamatan Sungai Pinang telah memasuki kondisi kurang sehat, terutama di area muara sungai. *Random Forest* diidentifikasi sebagai model terbaik karena interpretasi *feature importance*-nya, yang didukung oleh analisis SHAP, paling relevan dengan pemahaman ekologis. Temuan ini dapat menjadi dasar untuk pengelolaan sumber daya perairan yang lebih efektif.

Kata Kunci: Indeks Diversitas, Machine Learning, Random Forest, Sungai Ogan, Status Perairan

PENDAHULUAN

Perubahan iklim telah menjadi tantangan bagi kelestarian ekosistem perairan. Dampak dari perubahan ini menuntut pendekatan inovatif dalam pengelolaan keanekaragaman hayati, terutama di perairan sungai. Dalam konteks ini, komunitas ikan memiliki peran penting sebagai indikator utama untuk mengevaluasi status ekologis perairan tersebut. Meskipun teknologi biomonitoring dan penginderaan jauh telah berkembang, kesenjangan data masih menjadi tantangan. Banyak bagian sungai yang sulit diakses, sehingga informasi mengenai struktur komunitas ikan dan responsnya terhadap perubahan lingkungan menjadi terbatas (Sharma, 2023; Meinam, et al., 2023; Gavioli, et al., 2024; Brown, et al., 2024; Corona and Hogue, 2025).

Metode statistik dan model matematika konvensional sering kesulitan menangkap kompleksitas ekologis dalam keanekaragaman hayati akuatik. Namun, pendekatan berbasis data, seperti analisis Bayesian dan *machine learning*, termasuk kecerdasan buatan (AI), telah muncul sebagai alternatif yang lebih efektif. Metode-metode ini mampu meningkatkan pemahaman kita tentang ekosistem kompleks dengan mengungkap perubahan dan interaksi dinamis yang sering terabaikan oleh pendekatan lama (Lüdtke and Pierce, 2023; Danovaro, 2024; Shapen, et al., 2024; Miller, et al., 2025).

Studi ini mengambil Sungai Ogan sebagai studi kasus. Sungai Ogan, sebagaimana banyak perairan tawar, menghadapi tantangan ekologis substansial, termasuk degradasi habitat dan perubahan struktur komunitas ikan akibat tekanan lingkungan. Oleh karena itu, penelitian ini bertujuan untuk membuat model prediksi *machine learning* guna menginformasikan upaya konservasi dan keputusan kebijakan yang efektif dalam pengelolaan sumber daya perairan di Sungai Ogan dengan menerapkannya.

METODE PENELITIAN

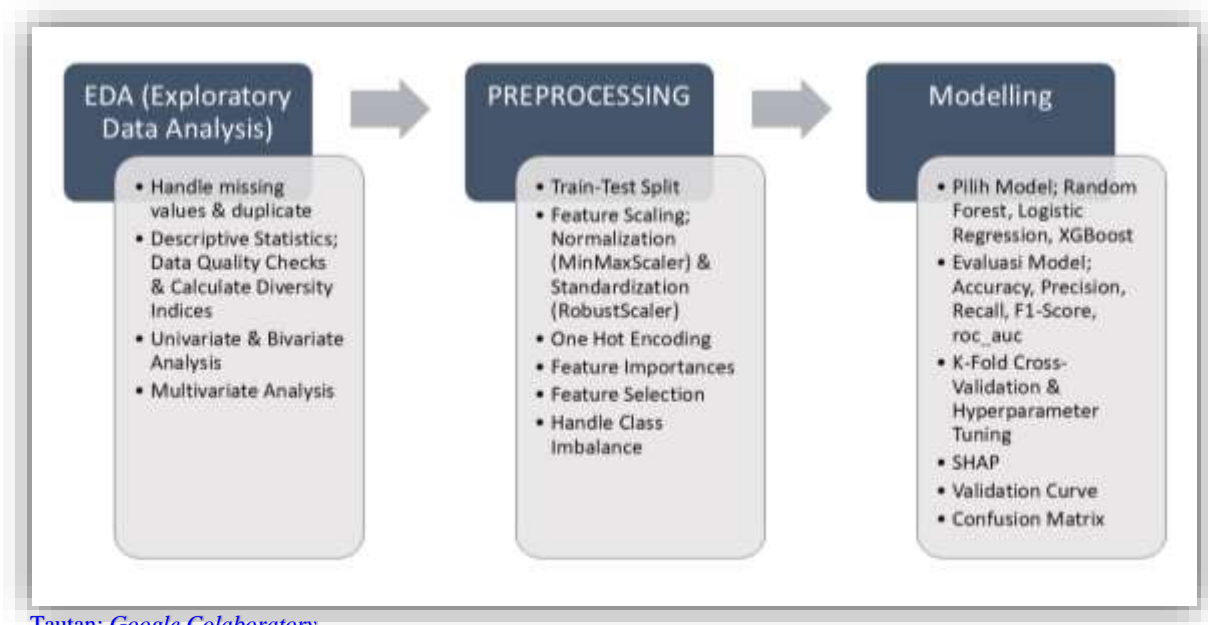
Pengumpulan data pada penelitian ini meliputi data spesies ikan yang tertangkap (jumlah spesies), kelimpahan ikan (jumlah individu) per alat tangkap dan per stasiun sampling, dan koordinat lokasi sampling (*longitude* dan *latitude*). Data ikan dikumpulkan dari bulan April hingga Juli di daerah aliran Sungai Ogan Kecamatan Sungai Pinang Kabupaten Ogan Ilir Sumatera Selatan melibatkan nelayan setempat. Pengamatan dan identifikasi ikan dilaksanakan di Laboratorium Perikanan Fakultas Perikanan, Universitas Islam Ogan Komering Ilir Kayuagung. Alat dan bahan yang digunakan adalah alat tangkap berupa jala dengan ukuran mata jaring 2-3 cm, tangkul dengan ukuran mata jaring 0,3 cm, jaring dengan ukuran mata jaring 2 cm, serok dengan ukuran mata jaring 0,3 cm, dan bubu dengan ukuran 1.5x2 cm, buku identifikasi, ikan sampel, dan formalin.

Lokasi pengambilan sampel ikan di tiga stasiun pengamatan (Gambar 1), yakni stasiun pertama yang terletak di bagian hulu Sungai Ogan di Desa Sungai Pinang, yang merupakan lokasi di sekitar penambangan pasir. Stasiun kedua bertempat Sungai Ogan di Desa Penyandingan, di wilayah yang terdapat kegiatan budidaya ikan sistem keramba, pencemaran rumah tangga yang bersumber dari limbah rumah tangga dan juga limbah pasar. Stasiun ketiga terletak di bagian hilir Sungai Ogan di Desa Penyandingan, yang merupakan muara aliran Sungai Ogan yang berada di Kecamatan Sungai Pinang.

Fokus penelitian ini adalah membuat model prediksi status perairan Sungai Ogan di lokasi pengamatan. Proses pembuatan model *machine learning* menggunakan *Python* dan *Jupyter Notebook* (*Google Colaboratory*) dimulai dengan menyusun *dataset*. *Dataset* dibuat dalam format CSV (*comma separated values*) yang kompatibel dengan *Python*, menggunakan data yang dikumpulkan melalui *Google Formulir*. Langkah-langkah detail dari seluruh proses ini dapat dilihat pada Gambar 2.



Gambar 1. Stasiun Sampling

Gambar 2. Alur Pembuatan Model *Machine Learning*

Tahap awal berfokus pada *Exploratory Data Analysis* (EDA). Tujuannya adalah untuk menghasilkan data yang bersih dan siap untuk pemodelan prediksi. Langkah-langkahnya dimulai dengan memeriksa data yang hilang dan duplikat, kemudian dilanjutkan dengan eksplorasi data melalui analisis univariat, bivariat, dan multivariat, serta penanganan data pencilan.

Tahap preprocessing adalah langkah krusial untuk mempersiapkan data. Tahap ini meliputi beberapa prosedur, dimulai dengan pembagian data menjadi data latih dan data uji. Selanjutnya, data mentah ditransformasi melalui proses normalisasi atau standardisasi untuk

fitur numerik, serta *One Hot Encoding* untuk fitur kategorikal. Untuk mengoptimalkan model, dilakukan *Feature Selection* berdasarkan hasil *Feature Importances* guna memastikan hanya fitur prediktor yang relevan yang digunakan. Terakhir, metode SMOTE (*Synthetic Minority Over-sampling Technique*) diterapkan untuk menyeimbangkan jumlah kelas, sehingga mengurangi potensi bias model.

Tahap terakhir adalah *modelling*. Tahap modelling berfokus pada pembangunan dan evaluasi model klasifikasi untuk memprediksi status perairan. Model yang dipilih terdiri dari tiga jenis: model non-linear *Random Forest* (RF) dan *XGBoost* (XGB), serta model linear *Logistic Regression* (LR) sebagai pembanding. RF dan XGB dipilih karena akurasi dan kemampuannya yang unggul dalam menangani data lingkungan yang kompleks, menjadikannya umum digunakan dalam prediksi berbasis data lingkungan selama lima tahun terakhir (Cao, et al., 2025). Sementara itu, LR digunakan karena koefisiennya mudah diinterpretasi, memberikan dasar yang jelas untuk membandingkan performa model non-linear.

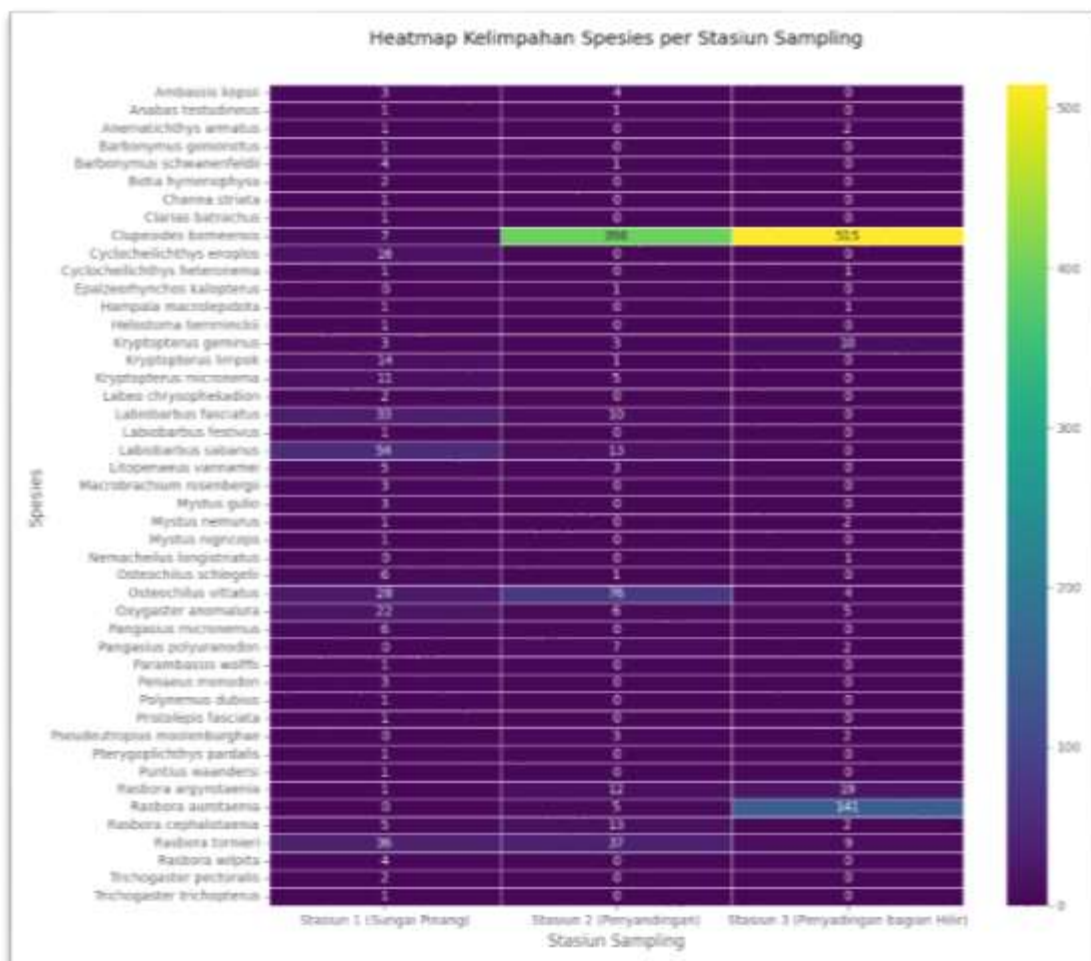
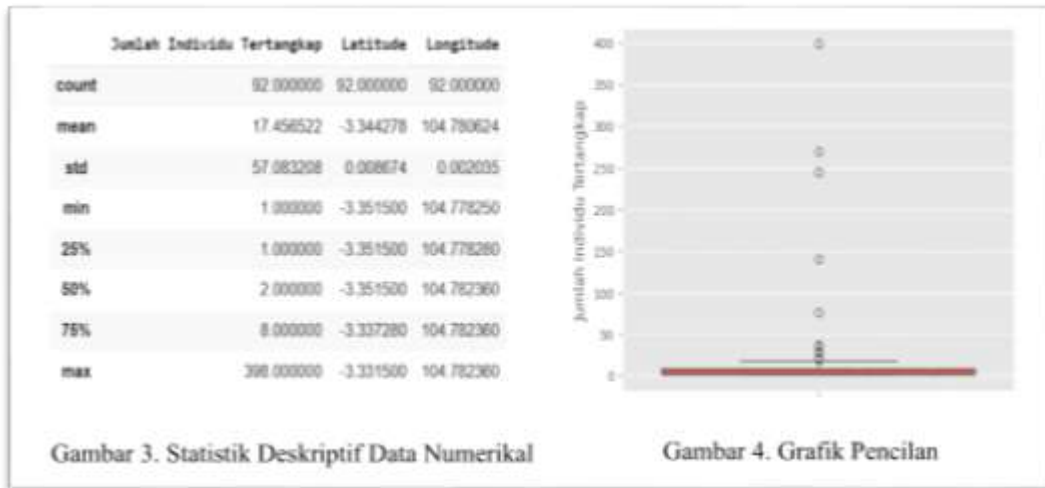
Evaluasi performa model dilakukan menggunakan beberapa metrik klasifikasi standar, seperti *classification_report*, *accuracy_score*, *precision_score*, *recall_score*, *f1_score*, dan *roc_auc_score*. Karena variabel target bersifat *multiclass*, *precision_score*, *recall_score*, dan *f1_score* dihitung menggunakan *average='weighted'* untuk mengatasi ketidakseimbangan kelas. Demikian pula, *roc_auc_score* dihitung dengan metode *weighted* dan *ovr*.

Untuk menganalisis perilaku model secara lebih mendalam, dilakukan beberapa langkah tambahan. Ini termasuk *K-Fold cross-validation & hyperparameter tuning* untuk mengoptimalkan model, SHAP (*SHapley Additive exPlanations*) untuk menjelaskan kontribusi setiap fitur, *validation curve* untuk menganalisis performa model terhadap parameter, dan *confusion matrix* untuk visualisasi hasil klasifikasi.

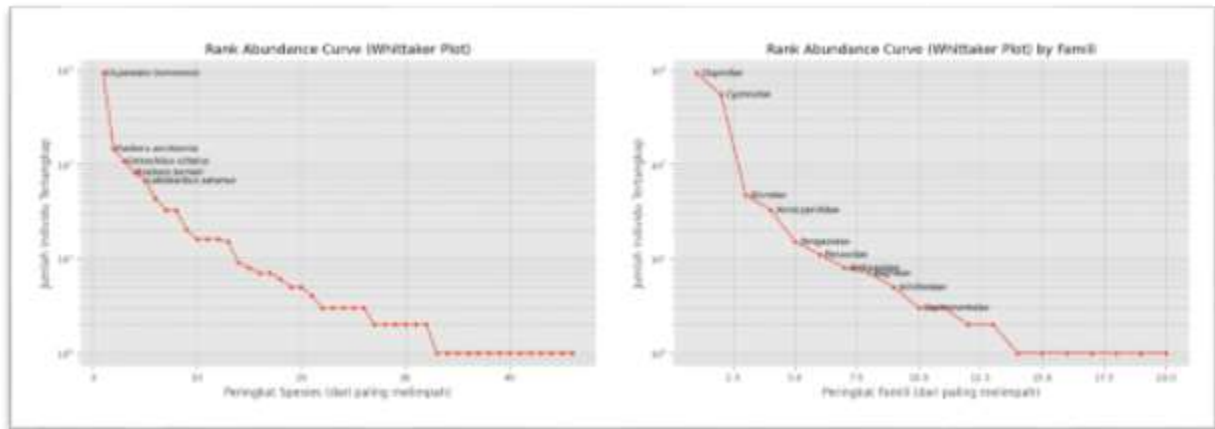
HASIL DAN PEMBAHASAN

Data statistik deskriptif numerikal menunjukkan adanya indikasi *outlier* (pencilan) pada fitur jumlah individu tertangkap (Gambar 3). Temuan ini diperkuat oleh grafik *boxplot* (Gambar 4) yang menunjukkan adanya pencilan tinggi. Analisis lebih lanjut (Gambar 5) mengungkapkan bahwa dominasi ini dipengaruhi oleh beberapa spesies tertentu. Di antaranya adalah *Clupeoides borneensis* yang banyak ditemukan di Stasiun 2 dan 3 Desa Penyandingan, *Rasbora aurotaenia* yang dominan di Stasiun 3 bagian hilir Desa Penyandingan, *Osteochilus vittatus* yang banyak ditemui di Stasiun 2 Desa Penyandingan, dan *Labiobarbus sabanus* yang dominan di Stasiun 1 Desa Sungai Pinang.

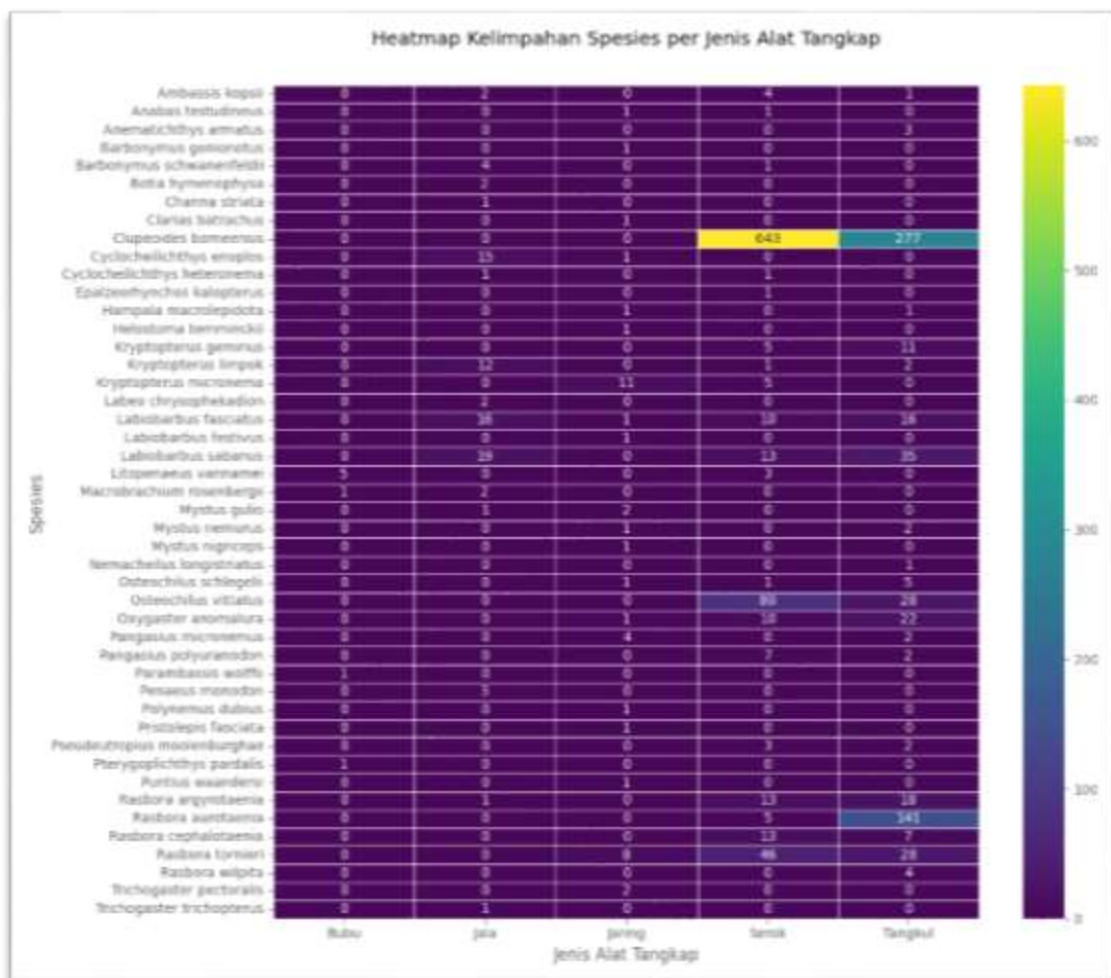
Menurut Iqbal et al. (2018), *Clupeoides borneensis* tersebar di perairan Kalimantan dan Asia Tenggara lainnya, termasuk di Daerah Aliran Sungai (DAS) Musi, yang mencakup Sungai Ogan. Adaptabilitas spesies ini dipengaruhi oleh sifat adaptif famili Clupeidae, yang mampu hidup di berbagai habitat (laut, payau, air tawar) selama jutaan tahun, sehingga menjadi kandidat ideal untuk memahami evolusi adaptif (Sebastian et al., 2022). Oleh karena itu, dominasi spesies ini di dua stasiun di Desa Penyandingan dapat menjadi bukti observasional lebih lanjut dari kemampuan adaptasi tersebut. Fakta ini diperkuat oleh *Rank Abundance Curve* (Gambar 6), yang menunjukkan bahwa *Clupeoides borneensis* berada pada urutan pertama dengan kelimpahan sekitar 10^3 . Selain itu, famili Cyprinidae, dengan spesies dominan seperti *Rasbora aurotaenia*, *Osteochilus vittatus*, dan *Labiobarbus sabanus*, juga menunjukkan pola yang sama. Famili Cyprinidae umumnya ditangkap menggunakan alat tangkap berbasis jaring, dalam hal ini jala, jaring, serok, dan tangkul sebagaimana ditunjukkan pada Gambar 7 (Iriansyah et al., 2021; Fatah, 2021).



Gambar 5. Heatmap Kelimpahan Spesies per Stasiun Sampling



Gambar 6. Rank Abundance Curve (Whittaker Plot)



Gambar 7. Heatmap Kelimpahan Spesies per Jenis Alat Tangkap

Statistik deskriptif data kategorikal (Gambar 8) menunjukkan bahwa dari 92 data observasi, sebagian besar dikumpulkan dari Stasiun 1 Desa Sungai Pinang (53 data). Serok menjadi alat tangkap yang paling banyak menghasilkan tangkapan, sementara spesies yang paling sering ditemukan adalah *Oxygaster anomalura* dari famili Xenocypridae (FishBase, 2025). Spesies ini tersebar di seluruh stasiun sampling dan tertangkap oleh lima jenis alat tangkap. Meskipun *Oxygaster anomalura* termasuk spesies yang paling sering ditemukan, famili yang paling dominan secara keseluruhan adalah Cyprinidae (42 data observasi).

Stasiun 1 Desa Sungai Pinang menunjukkan dominasinya terhadap data observasi karena beragamnya jenis alat tangkap yang digunakan di lokasi tersebut (Gambar 9). Penggunaan alat tangkap yang variatif ini berkorelasi dengan peningkatan variasi jenis ikan yang tertangkap (Gambar 10). Temuan ini konsisten dengan studi sebelumnya yang menunjukkan bahwa jenis alat tangkap yang berbeda secara signifikan memengaruhi jumlah dan variasi spesies ikan (Sullivan et al., 2015; Bhanja et al., 2024).

Serok (*Scoop Net/Dip Net*) dan juga tangkul (*Lift Net*), sebagaimana ditunjukkan pada Gambar 9 dan sempat disinggung sebelumnya, merupakan alat tangkap yang paling banyak menghasilkan tangkapan. Ini terjadi karena ukuran mata jaringnya relatif kecil (0,3 cm), sehingga spesies ikan yang masuk kelompok ikan kecil dan ikan sedang banyak tertangkap. Hal ini diperkuat dengan beberapa studi terdahulu, yang menemukan bahwa serok dan tangkul banyak digunakan di sungai untuk menangkap ikan kecil dan musiman, dan meskipun efektivitasnya cenderung lebih rendah dibanding alat modern, namun tetap penting secara ekonomi dan sosial, terutama untuk spesies tertentu dan pada musim tertentu (Mia et al., 2018; Hasmawati & Aras, 2023; Fishbase, 2025).

	Stasiun Sampling	Desa	Jenis Alat Tangkap	Spesies	Famili	Nama Lokal
count	92	92	92	92	92	92
unique	3	2	5	46	20	46
top	Stasiun 1 (Sungai Pinang)	Sungai Pinang	Serok	Oxygaster anomalura	Cyprinidae	Tortul/Samis
freq	53	53	27	5	42	5

Gambar 8. Statistik Deskriptif Data Kategorikal

Jenis Alat Tangkap		Spesies	
unique		count	
Stasiun Sampling		Jenis Alat Tangkap	
Stasiun 1 (Sungai Pinang)	4	Bubu	4
Stasiun 2 (Penyandingan)	1	Jala	15
Stasiun 3 (Penyandingan bagian Hilir)	2	Jaring	20
		Serok	27
		Tangkul	26

Gambar 9. Jumlah Alat Tangkap per Stasiun

Spesies		Spesies	
unique		count	
Stasiun Sampling		Stasiun Sampling	
Stasiun 1 (Sungai Pinang)	41	Stasiun 1 (Sungai Pinang)	53
Stasiun 2 (Penyandingan)	20	Stasiun 2 (Penyandingan)	20
Stasiun 3 (Penyandingan bagian Hilir)	15	Stasiun 3 (Penyandingan bagian Hilir)	19

Gambar 10. Variasi Spesies Ikan Tertangkap per Stasiun Sampling

Analisis selanjutnya berfokus pada indeks diversitas, yang meliputi Indeks Keanekaragaman Shannon-Wiener (H'), Indeks Kemerataan Pielou (J'), dan Indeks Dominansi Simpson (D). Perhitungan indeks-indeks ini dilakukan menggunakan Python, dan hasilnya

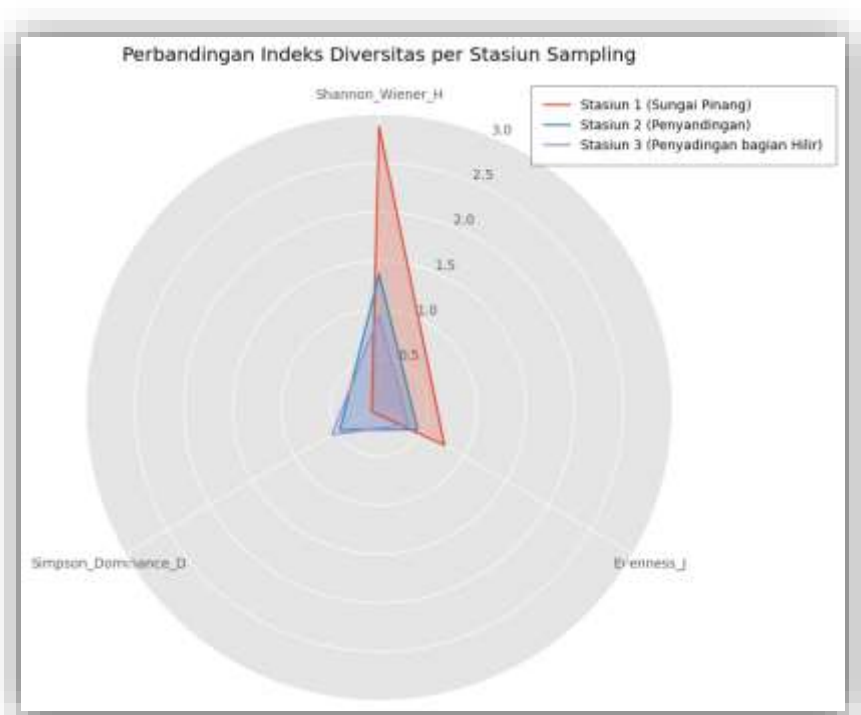
disajikan pada Gambar 11 dan 12. Setelah indeks diversitas diperoleh, dilakukan analisis multivariat untuk menguji korelasi antar-fitur numerik, yang divisualisasikan dalam bentuk *heatmap* (Gambar 13). Berdasarkan *heatmap*, ditemukan adanya korelasi antar-fitur seperti *latitude_summary*, *longitude_summary*, *total_individu*, *jumlah_spesies*, *Shannon_Wiener_H*, dan *Evenness_J*, *Simpson_Dominance_D*.

Status perairan ditentukan berdasarkan nilai indeks diversitas (Gambar 14). Interpretasi nilai Indeks Keanekaragaman Shannon-Wiener (H') mengikuti kriteria rendah (< 1), sedang ($1 \leq H' \leq 3$), dan tinggi (> 3) (Maretta et al., 2025). Untuk Indeks Kemerataan Pielou (J'), kategorisasinya adalah sangat rendah (< 0.25), rendah ($0.25-0.49$), sedang ($0.50-0.69$), tinggi ($0.70-0.89$), dan sangat tinggi (≥ 0.90) (Statology, 2025). Sementara itu, Indeks Dominansi Simpson (D) diklasifikasikan menjadi tidak ada dominansi ($< 0,5$) dan ada dominansi ($\geq 0,5$) (Febrian, 2022). Selanjutnya, status kesehatan perairan secara keseluruhan dikategorikan dengan mengombinasikan nilai dari ketiga indeks tersebut menjadi matriks penilaian, seperti yang ditunjukkan pada Tabel 1.

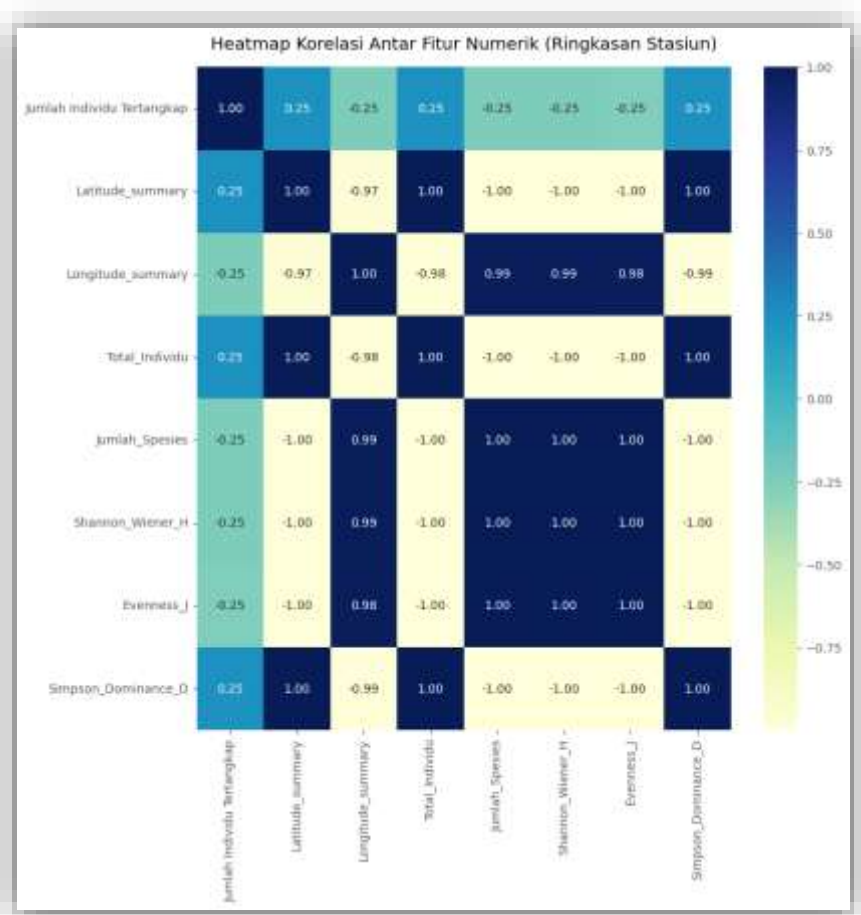
Analisis komparatif status perairan antar-stasiun menunjukkan variasi kondisi yang jelas. Stasiun 1 Desa Sungai Pinang memiliki status yang cukup sehat, dengan indeks keanekaragaman sedang, pemerataan baik, dan tanpa adanya dominansi. Stasiun 2 Desa Penyandingan, meskipun memiliki status cukup sehat dengan keanekaragaman sedang, menunjukkan indeks pemerataan yang rendah. Kondisi terburuk ditemukan di Stasiun 3 Desa Penyandingan (bagian hilir), yang memiliki status perairan kurang sehat dengan dominansi spesies yang jelas. Perbedaan kondisi ini dapat dikaitkan dengan lokasi spesifik setiap stasiun, di mana aktivitas antropogenik seperti penambangan pasir (Stasiun 1), budidaya ikan dan limbah domestik (Stasiun 2), serta lokasi muara sungai (Stasiun 3) diduga menjadi faktor penyebabnya. Beberapa penelitian terdahulu menguatkan temuan ini, dengan menunjukkan bahwa aktivitas antropogenik dapat merusak sumber daya air tawar dan berdampak negatif pada populasi ikan (Dudgeon, 2011; Kano et al., 2013; Liew et al., 2020; Yosieguspa et al., 2021, 2023).

	Shannon_Wiener_H	Evenness_J	Simpson_Dominance_D
Stasiun Sampling			
Stasiun 1 (Sungai Pinang)	2.870806	0.773058	0.088228
Stasiun 2 (Penyandingan)	1.362296	0.454746	0.461983
Stasiun 3 (Penyandingan bagian Hilir)	0.941292	0.347590	0.557317

Gambar 11. Indeks Diversitas



Gambar 12. Perbandingan Indeks Diversitas per Stasiun Sampling



Gambar 13. Heatmap Korelasi Antar Fitur Numerik

Indeks keanekaragaman per status dengan status perairan								
	Stasiun Sampling	Latitude	Longitude	Total_Individu	Jumlah_Species	Shannon_Wiener_H'	Evenness_J'	Simpson_Dominance_D
0	Stasiun 1 (Sungai Pramp)	-3.20160	104.76216	200	41	2.257000	0.773390	0.389225
1	Stasiun 2 (Pantai Bagan)	-3.23120	104.77820	600	20	1.362200	0.404740	0.401983
2	Stasiun 3 (Pantai Bagan Nelayan)	-3.23150	104.77820	710	15	0.941200	0.347500	0.337517

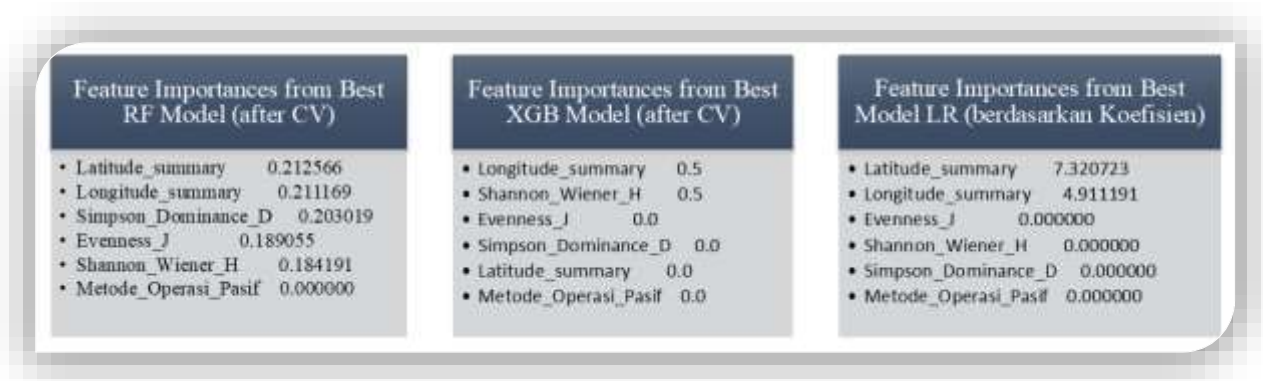
Gambar 14. Status Perairan

Tabel 1. Matriks Penilaian Status Perairan

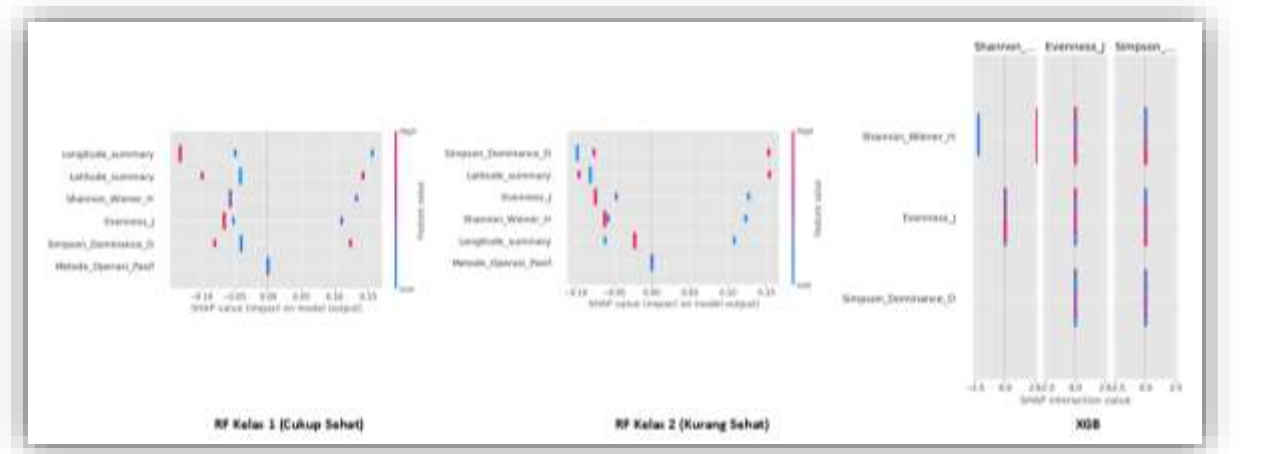
Indeks Keanekaragaman Shannon-Wiener (H')	Indeks Kemerataan Pielou (J')	Indeks Dominansi Simpson (D)	Status Perairan
Tinggi	Sangat tinggi / Tinggi	Tidak ada dominansi	Sangat Sehat
Sedang	Sangat tinggi / Tinggi/ Sedang = baik	Tidak ada dominansi	Cukup Sehat (Keanekaragaman Sedang, Kemerataan Baik, Tanpa Dominansi)
Sedang	Rendah/ Sangat rendah	Tidak ada dominansi	Cukup Sehat (Keanekaragaman Sedang, Kemerataan Rendah, Tanpa Dominansi)
Rendah	Rendah/ Sangat rendah	Ada dominansi	Kurang Sehat
Tinggi/ Sedang/ Rendah	Sangat Tinggi/ Tinggi/ Sedang/ Rendah/ Sangat rendah	Ada dominansi	Kritis (Ada Dominansi, Perlu Perhatian)
Lainnya	Lainnya	Lainnya	Status Campuran (Perlu Peninjauan Individual)

Setelah data di *preprocessing*, dilakukan penyesuaian untuk mengatasi ketidakseimbangan kelas pada variabel target (status perairan). Metode SMOTE berhasil menyeimbangkan jumlah data di setiap kelas, sehingga model dapat belajar dari setiap kategori status perairan secara merata. Hasil *feature selection* menunjukkan bahwa kombinasi antara indeks diversitas (H' , J' , D), fitur lokasi (*latitude_summary*, *longitude_summary*), dan jenis metode operasi (metode_operasi_pasif) memiliki pengaruh paling besar dalam memprediksi status perairan, sehingga fitur-fitur ini dipertahankan dalam model.

Evaluasi performa model pada *test set* menunjukkan bahwa *Random Forest*, *XGBoost*, dan *Logistic Regression* mencapai nilai *F1 score* dan metrik evaluasi lainnya sebesar 1.00. Kinerja sempurna ini mengindikasikan bahwa model-model tersebut mampu memprediksi setiap sampel pada *test set* dengan tepat. Performa luar biasa ini diduga kuat disebabkan oleh kombinasi fitur yang sangat unik dan diskriminatif pada setiap model: fitur lokasi dan indeks diversitas pada *Random Forest*; *longitude_summary* dan H' pada *XGBoost*; serta fitur lokasi pada *Logistic Regression*. Berdasarkan analisis, setiap kategori status perairan kemungkinan menempati "ruang" yang sangat berbeda dalam dimensi fitur-fitur tersebut (Gambar 15), membuat pemisahan kelas menjadi relatif mudah.

Gambar 15. *Feature Importance Model*

Analisis interpretatif menggunakan metode SHAP (*SHapley Additive exPlanations*) yang dilakukan pada model *Random Forest* dan *XGBoost* memberikan wawasan mendalam mengenai fitur-fitur yang paling berpengaruh dalam prediksi model. Seperti yang ditunjukkan pada Gambar 16, fitur lokasi (*longitude_summary* dan *latitude_summary*) serta indeks diversitas (*Shannon_Wiener_H*, *Evenness_J*, dan *Simpson_Dominance_D*) merupakan kontributor utama dalam menentukan status perairan. Temuan ini konsisten dengan hasil analisis multivariat dan *feature selection* yang menunjukkan bahwa status perairan sangat dipengaruhi oleh lokasi sampling dan karakteristik diversitas komunitas di setiap stasiun.

Gambar 16. *SHAP value Model*

Analisis terhadap indeks diversitas menunjukkan hubungan yang logis dengan kondisi ekologis. Karena model dilatih untuk menjelaskan probabilitas kondisi “Kurang Sehat”, maka nilai H’ dan J’ yang tinggi—yang secara ekologis merupakan indikator kondisi baik—akan menghasilkan *SHAP value* negatif. Hal ini menunjukkan bahwa kedua faktor tersebut menurunkan probabilitas suatu stasiun diklasifikasikan sebagai “Kurang Sehat”. Sebaliknya, nilai D yang tinggi (menunjukkan adanya dominansi, yang secara ekologis kurang baik) akan menghasilkan *SHAP value* positif, yang meningkatkan probabilitas model untuk memprediksi kondisi “Kurang Sehat”.

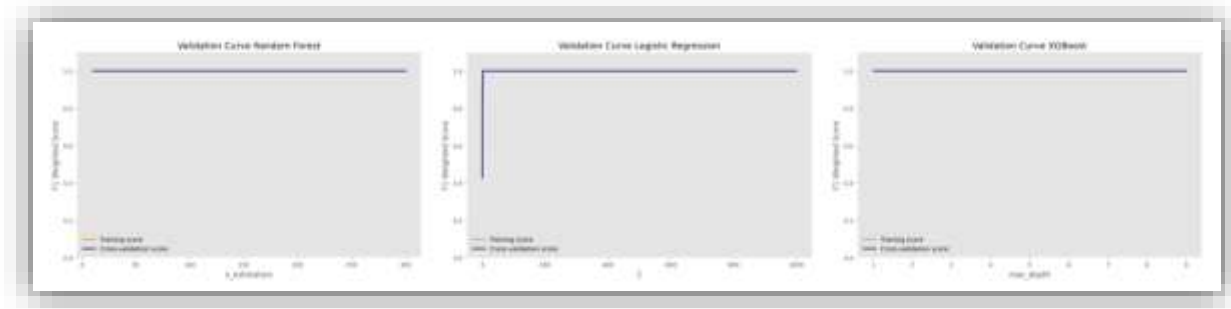
Pola interpretasi serupa juga ditemukan pada model lain seperti *XGBoost*, yang menunjukkan konsistensi temuan antar-model. Secara keseluruhan, temuan ini tidak hanya memvalidasi pemahaman ekologis terhadap indeks diversitas, tetapi juga menggarisbawahi pentingnya menjaga kondisi indeks diversitas yang baik dalam pengelolaan perairan.

Analisis menggunakan *validation curve* dilakukan untuk mengevaluasi performa model terhadap parameter-parameternya. Berdasarkan gambar 17, pada model *Random Forest* dan *XGBoost*, garis *training score* (garis oranye) dan *cross-validation score* (garis biru) sama-sama

berada di 1.00. Ini menunjukkan bahwa model tidak hanya mampu memprediksi data latih dengan sempurna, tetapi juga dapat menggeneralisasi dengan sempurna pada data yang belum pernah dilihat sebelumnya. Kondisi ini adalah bukti kuat bahwa model tidak mengalami *overfitting*, melainkan menemukan pola yang sangat jelas dan mudah dipisahkan di dalam data.

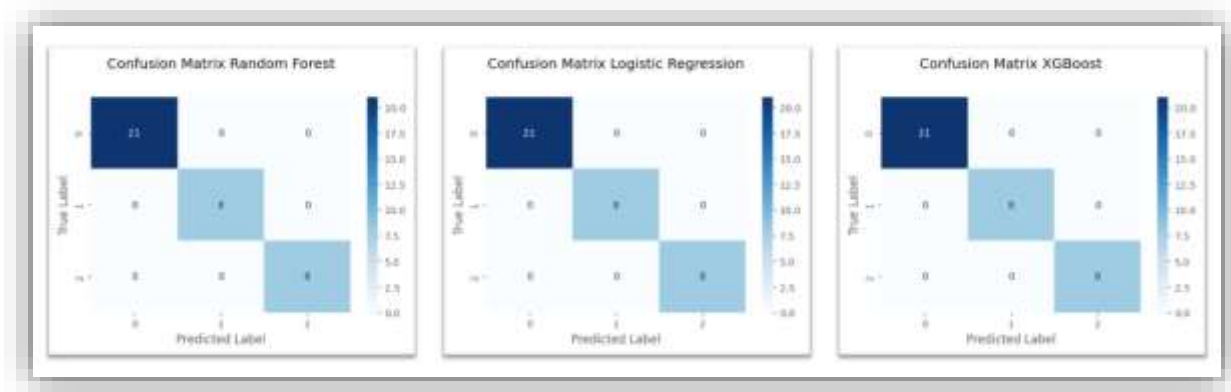
Pada model *Logistic Regression*, plot juga menunjukkan bahwa *training score* dan *cross-validation score* berada di angka 1.00 di seluruh rentang nilai parameter C. Fakta bahwa model yang relatif sederhana seperti *Logistic Regression* dapat mencapai performa sempurna menegaskan bahwa masalah klasifikasi ini, dengan fitur-fitur yang ada, secara fundamental sangat mudah untuk dipecahkan.

Semua bukti ini mengarah pada satu kesimpulan: fitur-fitur yang paling berpengaruh, terutama lokasi dan indeks diversitas yang diidentifikasi oleh SHAP, sangat efektif dan mampu memisahkan kelas-kelas dengan sempurna.



Gambar 17. Validation Curve Model

Analisis terakhir dilakukan dengan menggunakan confusion matrix untuk memvisualisasikan hasil klasifikasi model. Berdasarkan Gambar 18, terlihat bahwa model Random Forest, XGBoost, dan Logistic Regression berhasil memprediksi semua sampel dengan sempurna. Pada kelas 0, terdapat 21 sampel yang diprediksi dengan benar; pada kelas 1, terdapat 8 sampel yang diprediksi dengan benar; dan pada kelas 2, terdapat 8 sampel yang diprediksi dengan benar. Semua sel di luar diagonal utama bernilai nol, yang mengindikasikan bahwa ketiga model tidak melakukan satu pun kesalahan prediksi. Akurasi 1.00 pada data uji ini sangat konsisten dengan temuan sebelumnya dari validation curve dan menggarisbawahi bahwa fitur-fitur yang digunakan sangat efektif dan mampu memisahkan kelas-kelas secara optimal.



Gambar 18. Confusion Matrix Model

KESIMPULAN

Rangkaian analisis yang dilakukan, menuju kepada suatu kesimpulan, bahwa kondisi perairan di sepanjang Sungai Ogan di Kecamatan Sungai Pinang sudah memasuki kondisi kurang sehat, terlihat dari status perairan di muara sungai (stasiun 3). Artinya, perlu dilakukan upaya pencegahan dan perbaikan agar kondisi ini tidak berlanjut. Misalkan dengan membatasi kegiatan antropogenik yang berakibat negatif ke lingkungan perairan. Model prediksi yang dipilih, yakni RF, XGB, dan LR, menunjukkan performa yang sangat baik dengan akurasi 1.00 pada data uji. Namun, di antara ketiganya, model *Random Forest* (RF) dianggap yang paling tepat untuk studi ini. Hal ini didasarkan pada analisis interpretatif SHAP, yang secara jelas menunjukkan bahwa fitur lokasi (*longitude* dan *latitude*) dan indeks diversitas (H' , J' , dan D) merupakan kontributor utama dalam prediksi model. Karena fitur-fitur ini sangat relevan dengan tujuan ekologis studi, interpretasi yang dihasilkan oleh model *Random Forest* dianggap paling komprehensif dan sejalan dengan pemahaman saintifik tentang kondisi perairan.

DAFTAR PUSTAKA

- Bhanja, A., Payra, P., & Mandal, B. (2024). A Study on the Selectivity of Different Fishing Gear, *Ind. J. Pure App. Biosci.* 12(2), 8-19. doi: <http://dx.doi.org/10.18782/2582-2845.9072>.
- Brown, T. M., O'Connor, J. & Genner M. J. (2024). Climate warming drives population trajectories of freshwater fish. *Proc. Natl. Acad. Sci. U.S.A.*, 121 (51). <https://doi.org/10.1073/pnas.2410355121>.
- Cao, Y., Schartel, T., Houghton, D. C., Ross, J., & Infante, D. (2025). Evaluating machine learning algorithms for accuracy, stability, and among-predictors discriminability in modeling species-richness across ten datasets. *Ecological Informatics*, 90, 103323. <https://doi.org/10.1016/j.ecoinf.2025.103323>.
- Corona C. R. & Hogue, T. S. (2025). Machine Learning in Stream/River Water Temperature Modeling: A Review and Metrics for Evaluation. *Hydrol. Earth Syst. Sci.*, 29(12), 2521-2549. <https://doi.org/10.5194/hess-29-2521-2025>.
- Danovaro, R. (2024). Understanding Marine Biodiversity Patterns and Drivers: The Fall of Icarus. *Marine Ecology*, 00, e12814. <https://doi.org/10.1111/maec.12814>.
- Dudgeon, D. (2011). Asian river fishes in the Anthropocene: threats and Conservation Challenges in an Era of Rapid Environmental Change. *Journal of Fish Biology*, 79(6), 1487–1524. <https://doi.org/10.1111/J.1095-8649.2011.03086.X>.
- Fatah, K. (2021). Fish composition of Gillnet Cathces in Wadaslintang Reservoir, Central Java. *IOP Conference Series: Earth and Environmental Science*, 860. <https://doi.org/10.1088/1755-1315/860/1/012109>.
- Febrian, I., Nursaadah E., & Karyadi B. (2022). Analisis Indeks Keanekaragaman, Keragaman, dan Dominansi Ikan di Sungai Aur Lemau Kabupaten Bengkulu Tengah. *Bioscientist : Jurnal Ilmiah Biologi*, 10(2), 600-612. <https://doi.org/10.33394/bioscientist.v10i2.5056>.
- FishBase. (2025). *Clupeoides borneensis*. Diakses pada 9 Agustus 2025, dari https://www.fishbase.se/summary/Clupeoides_borneensis.html.
- FishBase. (2025). *Osteochilus vittatus*. Diakses pada 9 Agustus 2025, dari https://www.fishbase.se/summary/Osteochilus_vittatus.html.
- FishBase. (2025). *Oxygaster anomalura*. Diakses pada 9 Agustus 2025, dari https://www.fishbase.se/summary/Oxygaster_anomalura.html.
- FishBase. (2025). *Rasbora aurotaenia*. Diakses pada 9 Agustus 2025, dari https://www.fishbase.se/summary/Rasbora_aurotaenia.html.

- FishBase. (2025). *Rasbora tornieri*. Diakses pada 9 Agustus 2025, dari https://www.fishbase.se/summary/Rasbora_tornieri.html.
- Gavioli, A., Castaldelli, G., Trasforini, S., Puzzi, C., Gervasio, M. P., Granata, T., Colombo, D. & Soana, E. (2024). Global Warming and Fish Diversity Changes in the Po River (Northern Italy). *Environments*, 11(10), 226. <https://doi.org/10.3390/environments11100226>.
- Hasmawati, H., & Aras, M. (2023). Oceanographic Parameters of the Fishing Ground and Catch Results of Sodo' (Scope Net) in Mallusetasi Waters, Barru Regency. *Journal of Agriculture*, 1(3), 136-143. <https://doi.org/10.47709/joa.v1i03.2422>.
- Iqbal, M., Yustian, I., Setiawan, A. & Setiawan, D. (2018). *Ikan-Ikan di Sungai Musi dan Pesisir Timur Sumatera Selatan*. Palembang: Yayasan Kelompok Pengamat Burung Spirit of South Sumatra.
- Iriansyah., Aminah, S., & Gani, A. (2021). Selectivity of Rengge Hanyut Tools (Drift Gill Net) in the Sungai Batang Village of Martapura (Banjar District, South Kalimantan, Indonesia). *Russian Journal of Agricultural and Socio-Economic Sciences*, 4(112), 160-166. <https://doi.org/10.18551/rjoas.2021-04.20>.
- Kano, Y., Adnan, M. S., Grudpan, C., Grudpan, J., Magtoon, W., Musikasinthorn, P., Natori, Y., Ottomanski, S., Praxaysonbath, B., Phongsa, K., Rangsiruji, A., Shibukawa, K., Shimatani, Y., So, N., Suvarnaraksha, A., Thach, P., Nguyen Thanh, P., Tran, D. D., Utsugi, K., & Yamashita, T. (2013). An Online Database on Freshwater Fish Diversity and Distribution in Mainland Southeast Asia. *Ichthyological Research*, 60(3), 293–295. <https://doi.org/10.1007/S10228-013-0349-8>.
- Liew, J. H., Lim, R. B., Low, B. W., Mowe, M. A., Ng, T. H., Zeng, Y. W., & Yeo, D. C. (2020). *Tropical Freshwater Ecosystems, Biota and Anthropogenic Activities with Reference to South-East Asia*. In: Woo, P. T. K., Leong, J., Buchmann, K. (eds) Climate change and infectious fish diseases (pp. 19-43). Wallingford UK: CABI. <https://doi.org/10.1079/9781789243277.0019>.
- Lüdtke, S., & Pierce, M. (2023). Towards Machine Learning-based Fish Stock Assessment. *arXiv.Org*, abs/2308.03403. <https://doi.org/10.48550/arxiv.2308.03403>.
- Meinam, M., Singh, Y. J., Bharati, H. & Meinam, T. (2023). Importance of Fish Biodiversity Conservation and Management. *IJSRA*, 09(02), 387-391. <https://doi.org/10.30574/ijrsra.2023.9.2.0513>.
- Mia, Md. R., Arif, A. S. Md., Das, P. R., Islam, R., & Hossen, Md. A. (2018). Fishing Gears and Crafts of the Surma River in Sylhet District, Bangladesh. *International Journal of Fisheries and Aquatic Studies*, 6(4), 399–407. <http://www.fisheriesjournal.com/archives/2018/vol6issue4/PartE/6-3-32-172.pdf>.
- Miller, T., Michoński, G., Durlík, I., Kozłowska, P., & Biczak, P. (2025). Artificial Intelligence in Aquatic Biodiversity Research: A PRISMA-Based Systematic Review. *Biology*, 14(5), 520. <https://doi.org/10.3390/biology14050520>.
- Sebastian, W., Sukumaran, S., & Gopalakrishnan, A. (2022). Comparative mitogenomics of Clupeoid fish provides insights into the adaptive evolution of mitochondrial oxidative phosphorylation (OXPHOS) genes and codon usage in the heterogeneous habitats. *Heredity*, 128(4), 236–249. <https://doi.org/10.1038/s41437-022-00519-z>.
- Shapen, K. M., Akhmedov, A., & Rakhymova, A. T. (2024). Application of bayesian analysis for comprehensive assessment of the state of fish resources in freshwater ecosystems. *Habarşy - Abaj Atyndağy Almaty Universiteti. Fizika-Matematika Seriâsy*, 88(4). <https://doi.org/10.51889/2959-5894.2024.88.4.007>.
- Sharma, S.K. (2022). *Impact of Global Warming on Changing Pattern of Biodiversity and Fish Production in Inland Open Waters*. In: Sinha, A., Kumar, S., Kumari, K. (eds) Outlook

- of Climate Change and Fish Nutrition (pp. 49-61). Singapore: Springer.
https://doi.org/10.1007/978-981-19-5500-6_5.
- Statology, (2025). *How to Calculate and Interpret Pielou's Evenness Index*. Diakses pada 26 Juli 2025, dari [How to Calculate and Interpret Pielou's Evenness Index](#)
- Sullivan, C. J., Coulter, D. P., Feiner, Z. S., Donabauer, S. B., & Höök, T. O. (2015). Influences of Gear Type and Analytical Methodology on Fish Assemblage Characterisations in Temperate Lakes. *Fisheries Management and Ecology*, 22(5), 388–399.
<https://doi.org/10.1111/FME.12138>.
- Yosieguspa, Y., Fahleny, R., & Yuliani, Y. (2021). Analisis Mutu Air Akibat Aktivitas Penambang Pasir dengan Metode Storet di Sungai SP Padang OKI. *Jurnal Akuakultur Rawa Indonesia*, 9(1), 22–29. <https://doi.org/10.36706/jari.v9i1.14285>.
- Yosieguspa, Y., Fahleny, R., & Yuliani, Y. (2023). Study of Water Quality Effect Due to Floating Net Cages using the Storet Method in the River SP Padang OKI. *International Journal of Progresive Science and Technologi*, 38(02), 124-129.
<http://dx.doi.org/10.52155/ijpsat.v38.2.5247>.