

## Pemanfaatan *Random Forest* Dalam Pemetaan Status Trofik Di Danau Semayang Menggunakan Citra Sentinel-2 MSI

Muhammad Dika Oktoberdinata<sup>1\*</sup>, Pramaditya Wicaksono<sup>2</sup>, Sudaryatno<sup>2</sup>

<sup>1</sup>Magister Penginderaan Jauh, Fakultas Geografi, Universitas Gadjah Mada

<sup>2</sup>Departemen Sains Informasi Geografi, Fakultas Geografi, Universitas Gadjah Mada

\*Korespondensi: dika@mail.ugm.ac.id

### Abstrak

Pemantauan status trofik perairan menjadi aspek penting dalam menjaga kualitas ekosistem danau secara berkelanjutan. Penelitian ini bertujuan untuk memetakan dan menganalisis distribusi spasial status trofik Danau Semayang menggunakan metode Regresi *Random Forest* berdasarkan Indeks Carlson. Data yang digunakan meliputi hasil pengukuran parameter kualitas air (Klorofil-a, Total Fosfor, dan kedalaman cakram Secchi) pada Oktober 2024 serta citra Sentinel-2 MSI. Analisis dilakukan dengan membangun model hubungan antara nilai reflektansi band citra dan parameter kualitas air untuk memperoleh nilai Trophic State Index (TSI). Hasil penelitian menunjukkan bahwa model *Random Forest* mampu memprediksi nilai TSI dengan tingkat akurasi yang bervariasi, dengan nilai  $R^2$  masing-masing sebesar 0,3709 untuk kedalaman cakram Secchi, 0,2877 untuk Klorofil-a, dan 0,0277 untuk Total Fosfor. Secara spasial, Danau Semayang didominasi oleh kondisi mesotrofik dengan kecenderungan eutrof ringan, terutama di bagian utara dan timur laut danau. Temuan ini menunjukkan bahwa penggunaan citra Sentinel-2 MSI memiliki potensi besar untuk mendukung pemantauan status trofik perairan secara efisien dan berkelanjutan, khususnya pada ekosistem danau paparan banjir yang dinamis seperti Danau Semayang.

**Kata Kunci:** Sentinel-2 MSI, *Random Forest Regression*, *Trophic State Index* Carlson, Danau Semayang.

### Abstract

*Monitoring the trophic status of aquatic ecosystems is essential for maintaining lake water quality in a sustainable manner. This study aims to map and analyze the spatial distribution of the trophic status of Lake Semayang using the Random Forest Regression method based on the Carlson Trophic State Index (TSI). The data used include in-situ measurements of water quality parameters (Chlorophyll-a, Total Phosphorus, and Secchi Disk Depth) collected in October 2024, as well as Sentinel-2 MSI imagery. The analysis was conducted by modeling the relationship between the reflectance values of the satellite bands and the measured water quality parameters to estimate TSI values. The results indicate that the Random Forest model produced varying levels of accuracy, with  $R^2$  values of 0.3709 for Secchi Disk Depth, 0.2877 for Chlorophyll-a, and 0.0277 for Total Phosphorus. Spatially, Lake Semayang is dominated by mesotrophic conditions with a tendency toward light eutrophy, particularly in the northern and northeastern parts of the lake. These findings demonstrate that Sentinel-2 MSI imagery has great potential to support efficient and sustainable monitoring of trophic status, especially in dynamic floodplain lake ecosystems such as Lake Semayang.*

**Keyword:** Sentinel-2 MSI, *Random Forest Regression*, *Trophic State Index* Carlson, Lake Semayang.

## Pendahuluan

Danau merupakan bagian penting dari ekosistem perairan darat (limnik) dan memiliki peran strategis dalam mendukung fungsi ekologi, ekonomi, dan sosial masyarakat (Darajati dkk, 2016). Indonesia memiliki sekitar 5.807 danau yang tersebar di seluruh Indonesia (Lutfhi, 2020). Salah satu kawasan danau yang ditetapkan sebagai danau prioritas nasional adalah kawasan Danau Kasakade Mahakam (DKM), yang terletak di bagian tengah Daerah Aliran Sungai Mahakam (DAS Mahakam). DKM merupakan kumpulan danau yang terdiri dari 20 danau dengan tiga danau terbesarnya adalah Danau Jempang (15000 ha), Danau Semayang (13000 ha), dan Danau Melintang (11000 ha) (Panitia Kerja Penyusunan Rencana Pengelolaan Danau Kaskade Mahakam, 2019).

Danau Semayang menjadi salah satu danau yang paling banyak disorot karena mengalami permasalahan lingkungan seperti pendangkalan, dan pertumbuhan gulma air yang masif sehingga dapat menutupi permukaan danau (Mislan dkk, 2022). Masalah tersebut berkaitan erat dengan peningkatan kandungan nutrient di perairan, seperti fosfor dan nitrogen yang akhirnya dapat memicu terjadinya proses eutrofikasi (Ariyanto dkk, 2014; Asdak, 2007; Mahmud & Achmad, 2020). Salah satu metode yang paling umum digunakan untuk mengukur tingkat eutrofikasi adalah menggunakan Indeks Status Trofik Carlson (*Carlson Trophic State Index / CTSI*), yang memanfaatkan tiga parameter kualitas air seperti konsentrasi Klorofil-a, Total Fosfor, dan kejernihan perairan yang ditunjukkan dengan kedalaman Secchi Disk (Carlson, 1977; Klippel dkk, 2020).

Umumnya, ketiga parameter tersebut diperoleh melalui pengambilan sampel lapangan dan analisis laboratorium. Sehingga, hasil yang diperoleh lebih akurat. Namun, pendekatan tersebut memiliki keterbatasan dari segi waktu, biaya, tenaga, dan cakupan spasial yang terbatas (Kutser dkk, 2020; Meng dkk, 2023). Dalam konteks ini, penginderaan jauh dapat menjadi alternatif dalam membantu menjawab keterbatasan yang ada agar lebih efisien dalam memantau kualitas air secara luas dan konsisten dalam jangka waktu tertentu (Kutser dkk, 2020). Penelitian terdahulu menggunakan citra satelit dalam mengestimasi parameter kualitas air dan status trofik dengan pendekatan statistik maupun *machine learning*, seperti menggunakan regresi linier yang dilakukan oleh Priadi & Sudaryatno (2016), Rivani & Wicaksono (2018), dan Regresi *Random Forest* yang dilakukan oleh Amanatullah (2023).

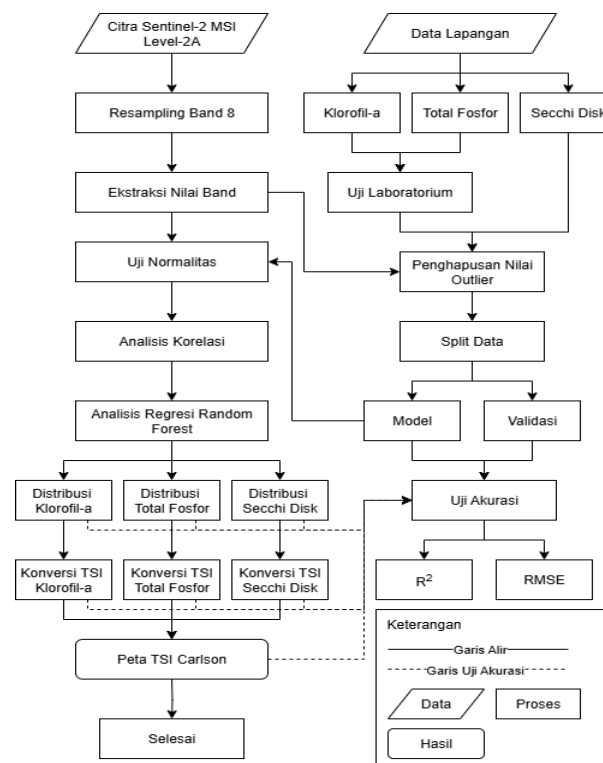
Namun, masih terdapat kekosongan studi khususnya dalam mengestimasi status trofik berbasis penginderaan jauh pada danau paparan banjir, seperti Danau Semayang yang memiliki karakteristik kompleks dan berfluktuasi secara musiman. Berdasarkan hal tersebut, tujuan dari penelitian ini adalah untuk mengestimasi dan memetakan status trofik Pemanfaatan *Random Forest* Dalam Pemetaan Status Trofik...  
Muhammad Dika Oktoberdinata, Pramaditya Wicaksono, Sudaryatno

di Danau Semayang berdasarkan algoritma Regresi *Random Forest* menggunakan citra sentinel-2 MSI. Penelitian ini berfokus pada tahun 2024, saat kondisi perairan Danau Semayang mulai meningkat akibat pasang dari Sungai Mahakam.

## Metodelogi Penelitian

### Kerangka Penelitian

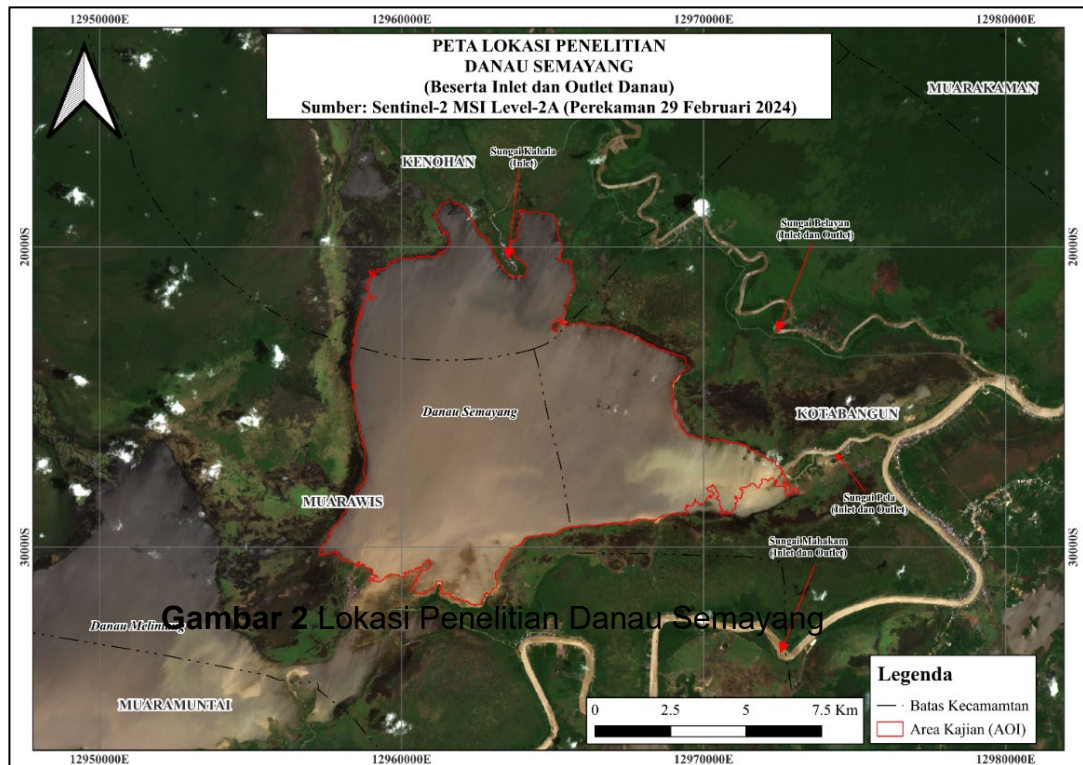
Diagram alir pada Gambar 1, menunjukkan tahapan penelitian status trofik perairan dengan menggunakan data citra Sentinel-2 MSI Level-2A dan data lapangan berupa Klorofil-a, Total Fosfor, dan Secchi Disk. Data citra terlebih dahulu dilakukan resampling Band 8 dan ekstraksi nilai band 2 hingga band 12, sedangkan data lapangan melalui proses uji laboratorium dan penghapusan nilai outlier. Selanjutnya, dilakukan uji normalitas untuk memastikan distribusi data, serta pembagian data menjadi set model dan validasi. Analisis korelasi digunakan untuk mengukur hubungan antara nilai reflektansi dengan parameter kualitas air, yang kemudian menjadi input pada analisis regresi *Random Forest* guna memprediksi distribusi Klorofil-a, Total Fosfor, dan kejernihan Secchi Disk. Hasil prediksi diverifikasi melalui uji akurasi menggunakan koefisien determinasi ( $R^2$ ) dan *Root Mean Square Error* (RMSE). Prediksi distribusi selanjutnya dikonversi ke indeks *Trophic State Index* (TSI) berdasarkan parameter Klorofil-a, Total Fosfor, dan Secchi Disk, yang kemudian disintesis menjadi peta TSI Carlson sebagai output akhir pemetaan status trofik perairan.



**Gambar 1** Diagram Alir Penelitian

## Lokasi Penelitian

Penelitian ini dilakukan di Danau Semayang yang terletak di Kabupaten Kutai Kartanegara Provinsi Kalimantan Timur. Secara geografis, Danau Semayang terletak dalam UTM Zona 50 M pada koordinat 433234,07 mE (*Easting*) hingga 448071.74 mE (*Easting*), dan 9981577.31 mN (*Northing*) hingga 9969692.11 mN (*Northing*). Secara administrasi danau ini berada di tiga kecamatan yaitu Kecamatan Kenohan, Kecamatan Muara Wis, dan Kecamatan Kota Bangun (Gambar 2).



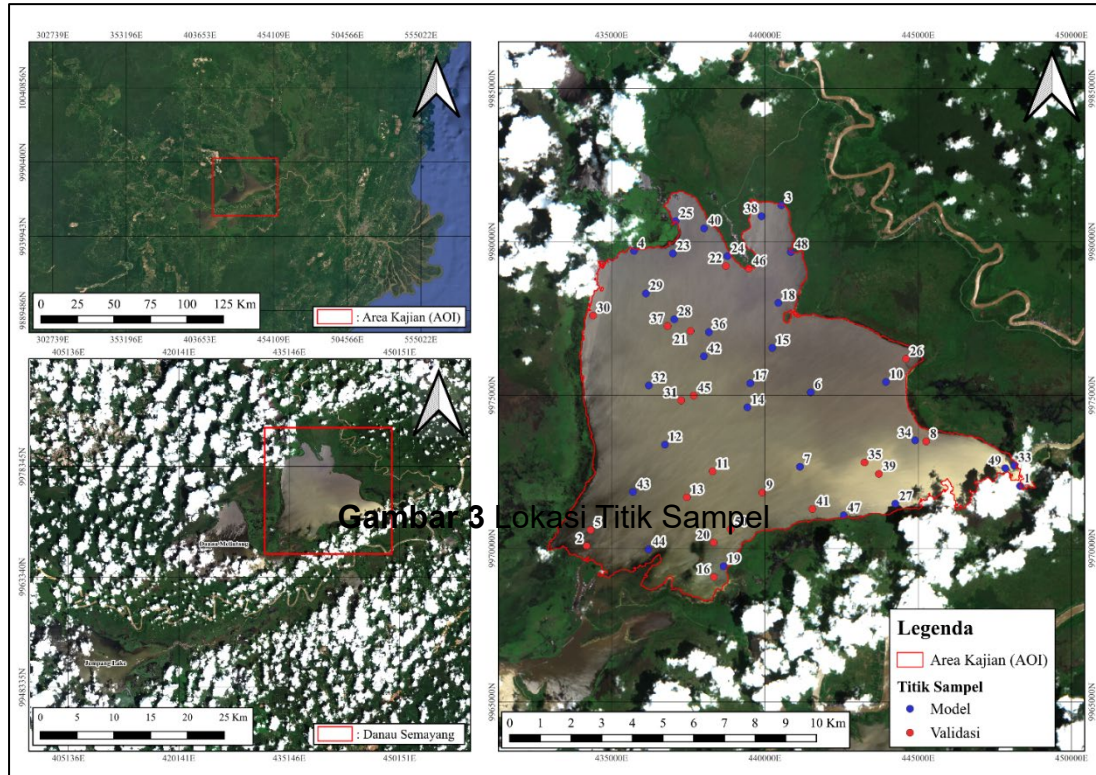
Danau semayang merupakan sebuah danau paparan banjir yang menjadi bagian dari wilayah retensi air dan penting bagi ekosistem lahan basah di DAS Mahakam. Kondisi morfologi yang datar menyebabkan perairan danau dapat mengalami perubahan ekstrem secara musiman dan dapat tergenang sepanjang tahun. Sungai utama yang berpengaruh terhadap kondisi perairan adalah Sungai Mahakam, dan anak-anak sungainya seperti Sungai Kahala, Sungai Belayan, dan Sungai Enggelam.

## Data dan Preprocessing

### Pengambilan data lapangan

Sebaran titik sampel ditentukan berdasarkan perbedaan nilai piksel hasil klasifikasi IsoDATA dan mempertimbangkan lokasi inlet dan outlet, serta penggunaan lahan disekitar danau. Citra yang digunakan dalam pembuatan peta sampling adalah citra Sentinel-2 MSI

dengan waktu perekaman 6 Oktober 2024. pemilihan tanggal perekaman tersebut karena ketersediaan data yang mendekati tanggal pengambilan data lapangan yaitu 13-14 Oktober 2024. Jumlah sampel lapangan sebanyak 50 titik dan tersebar diseluruh Danau (Gambar 3) dengan rincian 30 sampel model dan 20 sampel validasi. Adapun data yang diambil adalah data sampel air Klorofil-a dan Total Fosfor, dan kedalaman Secchi Disk.



Gambar 3 Lokasi Titik Sampel

### Citra Sentinel-2 MSI

Citra Sentinel-2 MSI yang digunakan untuk menganalisis status trofik di Danau Semayang menggunakan perekaman tanggal 5 November 2024. Dasar pemilihan citra tersebut karena kualitas citra pada tanggal tersebut memiliki tutupan awan yang sedikit. Selain itu, citra yang digunakan dalam penelitian ini adalah citra Level-2 yang merupakan citra sudah terkoreksi secara geometrik dan radiometrik sehingga bisa langsung digunakan untuk proses pemodelan.

Meskipun terdapat perbedaan waktu antara pengambilan data lapangan (13-14 Oktober 2024) dan citra Sentinel-2 yang digunakan (5 November 2024), pemilihan citra tersebut dipertimbangkan sebagai alternatif terbaik karena merupakan citra dengan tanggal terdekat yang memiliki tutupan awan paling sedikit pada area danau. Walaupun lama perbedaan waktu tersebut dapat mengganggu hasil analisis karena kondisi perairan yang begitu dinamis, citra Sentinel-2 perekaman 5 November 2024 dianggap masih representatif untuk menggambarkan kondisi perairan Danau Semayang pada periode pengukuran karena kondisi air yang masih dalam kondisi pasang.



Resolusi citra yang digunakan adalah 20 meter karena memiliki jumlah band yang bervariasi sehingga dapat digunakan untuk mencari dengan Band 8 yang dilakukan proses resampling dari 10 meter ke 20 meter. Informasi mengenai citra Sentinel-2 MSI Level-2A dapat dilihat pada Tabel 1.

**Tabel 1** Informasi Citra Sentinel-2 MSI Level-2A

No	Band - Keterangan	Panjang Gelombang Pertengahan ( $\mu\text{m}$ )	Panjang Gelombang ( $\mu\text{m}$ )	Resolusi Spasial (m)		
				10	20	60
1	Band 1 - <i>Coastal Aerosol</i>	0,443	0,433 – 0,453		√	√
2	Band 2 - <i>Blue</i>	0,490	0,458 – 0,529	√	√	√
3	Band 3 - <i>Green</i>	0,560	0,543 – 0,578	√	√	√
4	Band 4 - <i>Red</i>	0,665	0,650 – 0,680	√	√	√
5	Band 5 - <i>Vegetation Red Edge 1</i>	0,705	0,698 – 0,713		√	√
6	Band 6 - <i>Vegetation Red Edge 2</i>	0,740	0,733 – 0,748		√	√
7	Band 7 - <i>Vegetation Red Edge</i>	0,783	0,773 – 0,793		√	
8	Band 8 - <i>Near Infrared (NIR)</i>	0,842	0,785 – 0,900	√		√
9	Band 8A - <i>Vegetation Red Edge</i>	0,865	0,855 – 0,875		√	√
10	Band 9 - <i>Water Vapour</i>	0,945	0,935 – 0,955			√
11	Band 10 - <i>SWIR / Cirrus</i>	1,375	1,360 – 1,390		√	
12	Band 11 – <i>SWIR 1</i>	1,610	1,565 – 1,655		√	√
13	Band 12 – <i>SWIR 2</i>	2,190	2,100 – 2,280		√	√

Sumber: ESA (2015)

### Status Trofik Carlson

Dalam studi ini, status trofik diukur menggunakan metode TSI Carlson. Metode ini mempertimbangkan ketiga parameter status trofik, yaitu Klorofil-a, Total Fosfor, dan Secchi Disk. Adapun persamaannya adalah sebagai berikut:

$$TSI_{SD} = 60 - 14,4 \ln(SD) \quad (1)$$

$$TSI_{Chl-a} = 9,81 \ln(Chl - a) + 30,6 \quad (2)$$

$$TSI_{TP} = 14,2 \ln(TP) + 4,15 \quad (3)$$

$$CTSI = \frac{[TSI(SD) + TSI(Chl - a) + TSI(TP)]}{3} \quad (4)$$

Hasil perhitungan rata-rata nilai trofik dari TSI Carlson dapat dibandingkan dengan kategori status trofik Carlson seperti pada Tabel 2, sehingga dapat diketahui tingkat status trofik danau.

Tabel 2 Kelas Status Trofik Carlson

TSI Rata-rata	Status Trofik	Keterangan
< 30	Ultraoligotrof	Air jernih dengan kadar unsur hara sangat rendah.
30-40	Oligotrof	Air jernih dengan kadar unsur hara rendah.
40-50	Mesotrof	Kecerahan air sedang dengan kadar unsur hara sedang.
50-60	Eutrof Ringan	Penurunan kecerahan air dengan kadar unsur hara meningkat.
60-70	Eutrof Sedang	Marak alga ( <i>Microcystis</i> ) dengan kandungan hara tinggi.
70-80	Eutrof Berat	Marak alga dengan diikuti pertumbuhan gulma air secara cepat dan memiliki unsur hara sangat tinggi.
> 80	Hipereutrof	Marak alga dengan keadaan perairan dalam kondisi anoxia yang menyebabkan kematian ikan secara massal, selain itu juga kadar unsur hara amat sangat tinggi.

Sumber: Carlson, 1977

### Identifikasi dan Penghapusan *Outlier*

Dalam penelitian ini, pendekatan yang digunakan untuk mendeteksi *outlier* adalah metode *Interquartile Range* (IQR), yaitu dengan mengidentifikasi batas bawah dan batas atas dari distribusi data berdasarkan kuartil. Teknik ini dipilih karena lebih tahan terhadap pengaruh nilai ekstrim dibandingkan metode berbasis rata-rata atau simpangan baku. Selain itu, IQR tidak dibatasi oleh ketergantungan, asumsi, atau distribusi data Gaussian (Meng dkk., 2023), sehingga lebih fleksibel untuk berbagai jenis data lingkungan. Penghapusan *outlier* menjadi langkah penting karena data hidrologi kerap mengalami penyimpangan akibat kesalahan pengukuran atau faktor eksternal lainnya (Jeong dkk., 2017).

Identifikasi dan penghapusan *outlier* dilakukan terhadap 3 data parameter utama hasil pengukuran lapangan dan uji laboratorium. Ketiga data tersebut dipilih karena data tersebut mewakili kondisi aktual di lapangan dan digunakan sebagai variabel target dalam proses pemodelan regresi. Analisis ini dilakukan dengan memanfaatkan statistik deskriptif dan *boxplot* untuk visualisasi berdasarkan hasil IQR yang ditandai dengan adanya titik-titik terpisah pada *boxplot*. Cara kerja IQR adalah dengan menghitung selisih antara kuartil ketiga (Q3) dan kuartil pertama (Q1). Dimana suatu nilai yang dianggap *outlier* jika memenuhi kondisi seperti lebih besar dari  $Q3 + 1,5 \times IQR$  atau Lebih kecil dari  $Q1 - 1,5 \times IQR$ . Kondisi tersebut dapat dinyatakan dengan persamaan 5, 6, dan 7.

$$Outlier_{\epsilon} = (-\infty, E_{min}) \cup (E_{max}, +\infty) \quad (5)$$

$$E_{min} = Q1 - 1,5(Q3 - Q1) \quad (6)$$

$$E_{max} = Q3 + 1,5(Q3 - Q1) \quad (7)$$

Dimana Q1 dan Q3 merupakan kuartil pertama dan kuartil ketiga dari nilai estimasi yang telah diurutkan.

### Regresi Random Forest

Regresi *Random Forest* (RRF) adalah metode pembelajaran ensemble yang menggabungkan pohon keputusan dan keacakan untuk membentuk model prediktif yang kuat (Fu dkk, 2024). Teknik ini mengembangkan banyak pohon keputusan berdasarkan bagian acak dari data pelatihan dan subset variabel input yang dipilih secara acak (Islam dkk., 2023; Fu dkk., 2024), melalui dua Langkah utama yaitu Pembangunan pohon keputusan dan penerapan keacakan dalam setiap pohon (Liu dkk, 2019). Setiap pohon membuat prediksi secara independent, dan hasil akhir diperoleh dari rata-rata semua prediksi (Islam dkk, 2023). RRF efektif dalam menanangi data yang memiliki hubungan linier maupun non-linier, interaksi kompleks antar variabel, serta tahan terhadap overfitting. RRF juga coco diterapkan pada data berdimensi tinggi, baik data kategorikal maupun numerik, serta baik dalam mengahapi outlier dan data yang hilang (Zhou dkk, 2023). Dalam kajian hidrologi, RRF telah banyak diterapkan untuk menganalisis dan mengindentifikasi variabel prediktor penting dalam memahami proses perairan (Islam dkk, 2023). Dalam penelitian ini, RRF digunakan untuk memetakan distribusi parameter status trofik seperti Klorofil-a, Total Fosfor, dan Kedalaman Secchi Disk.

Model RRF dibangun dengan dua jenis konfigurasi yaitu menggunakan konfigurasi standar, dan konfigurasi hyperparamter tuning (Table 3). Kedua konfigurasi ini bertujuan untuk mencari konfigurasi terbaik dalam mengestimasi setiap parameter status trofik.

**Tabel 3** Konfigurasi Model RRF

Parameter	Keterangan	Tuning Standar	Hyperparameter Tuning
<b><i>n Estimator</i></b>	Jumlah pohon keputusan dalam <i>Random Forest</i> .	100	50, 100, 200, 500
<b><i>Max Depth</i></b>	Kedalaman maksimum setiap pohon keputusan.	10	None, 10, 20
<b><i>Random State</i></b>	Seed untuk menghasilkan hasil yang sama setiap kali model dijalankan.	42	42
<b><i>Min Samples Split</i></b>	Jumlah minimum sampel yang dibutuhkan untuk membagi node.	<i>Default</i>	2, 5
<b><i>Min Samples Leaf</i></b>	Jumlah minimum sampel di setiap <i>leaf</i> (simpul akhir).	<i>Default</i>	1, 2



Parameter	Keterangan	Tuning Standar	Hyperparameter Tuning
<b>Max Features</b>	Jumlah fitur maksimum yang dipertimbangkan di setiap split.	'sqrt'	'auto' atau 'sqrt' → memilih akar kuadrat dari total fitur.

### Uji Akurasi

Uji akurasi bertujuan untuk mengetahui seberapa baik hasil pemodelan yang telah dilakukan. Parameter yang diuji akurasinya adalah Klorofil-a, Total Fosfor, Secchi Disk, dan nilai TSI Carlson. Uji akurasi dilakukan dengan menggunakan beberapa matriks seperti Koefisien Determinasi ( $R^2$ ), dan *Root Means Square Error* (RMSE). Adapun persamaannya sebagai berikut:

$$R^2 = 1 - \frac{\sum_{i=1}^n (t_i - y_i)^2}{\sum_{i=1}^n (t_i - \bar{t})^2} \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (t_i - y_i)^2} \quad (9)$$

Dimana  $t$  adalah nilai aktual, dan  $y$  adalah hasil estimasi model, dan  $n$  adalah jumlah sampel.

## Hasil dan Pembahasan

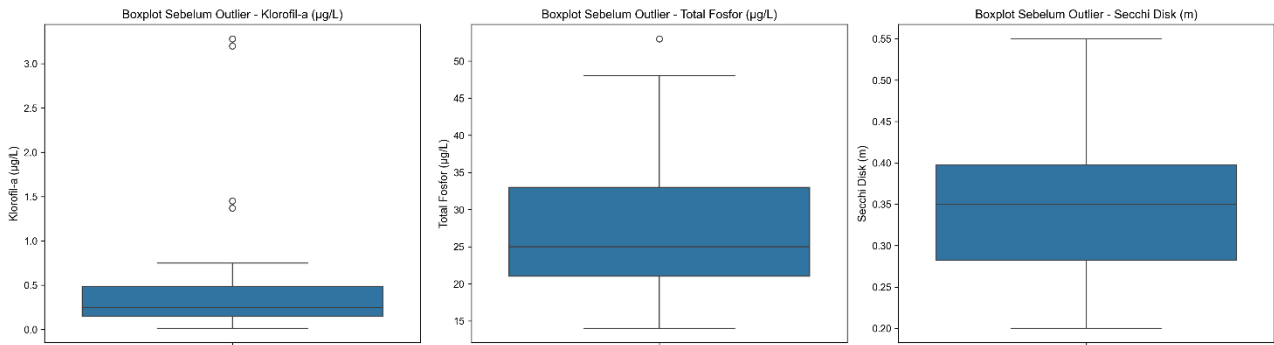
### Analisis Statistik

#### Sebelum dan Sesudah Penghapusan Outlier

Hasil statistik deskriptif terhadap 50 data sebelum dihapus outlier yang terdapat pada Tabel 4 menunjukkan rentang nilai yang lebar pada parameter klorofil-a dan total fosfor. Pada klorofil-a memiliki rentang dari 0,013 hingga 3,28  $\mu\text{g/L}$ . Sedangkan, total fosfor memiliki rentang dari 14 hingga 53  $\mu\text{g/L}$ . Lebar rentang nilai mengindikasikan adanya outlier pada data klorofil-a sebanyak 4 data dan total fosfor sebanyak 1 data yang ditunjukkan pada Gambar 4. Selain itu, Secchi disk memiliki rentang nilai yang tidak lebar dari 0,2 hingga 0,55 meter sehingga tidak terdapat outlier pada parameter tersebut.

**Tabel 4** Statistik Deskriptif Sebelum Hapus Outlier

Parameter	Jumlah	Mean	Std	Min	25%	50%	75%	Maks
<b>Klorofil-a</b>	50	0,449	0,648	0,013	0,1465	0,247	0,4855	3,28
<b>Total Fosfor</b>	50	27,32	9,460	14	21	25	33	53
<b>Secchi Disk</b>	50	0,343	0,080	0,2	0,2825	0,35	0,3975	0,55

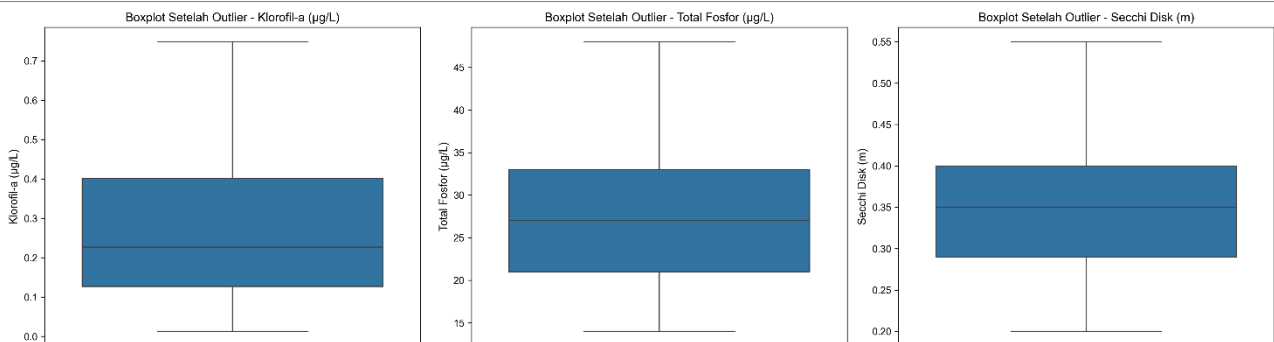


**Gambar 4** Boxplot Sebelum Hapus Outlier.

Berdasarkan hasil analisis IQR, teridentifikasi sebanyak 5 data outlier dari total 50 data observasi (Tabel 5). Data yang outlier terdapat pada variabel Klorofil-a dan Total Fosfor yang menunjukkan nilai ekstrim di luar rentang distribusi normalnya yang dapat dilihat pada boxplot Gambar 3. Distribusi data yang terdeteksi outlier terdapat pada data sampel nomor Setelah dilakukan penghapusan outlier, boxplot pada Gambar 5 menunjukkan distribusi yang lebih simetris dan tidak lagi terdapat nilai ekstrim. Hal tersebut mengindikasikan bahwa distribusi data menjadi lebih homogen dan mendekati distribusi normal.

**Tabel 5** Data yang terdeteksi Outlier

No	X	Y	Sampel	Jenis Sampel	Klorofil-a	Total Fosfor	Secchi Disk
1	116.465629	-0.170148024	TS3	M	1,37	22	0,33
2	116.474237	-0.225362973	TS6	M	0,214	53	0,26
3	116.43384	-0.184402037	TS23	M	1,45	16	0,44
4	116.533939	-0.246923994	TS33	M	3,28	16	0,23
5	116.490004	-0.246054959	TS35	V	3,2	24	0,22



**Gambar 5** Boxplot setelah hapus outlier

Hasil penghapusan membuat menyebabkan penurunan nilai maksimum dan standar deviasi, dan membuat nilai rata-rata menjadi lebih representatif terhadap keseluruhan data yang ditunjukkan pada Tabel 6. Selain itu, setelah dilakukan penghapusan outlier, data yang digunakan untuk analisis lebih lanjut berjumlah 45.

**Tabel 6** Statistik Deskriptif Sesudah Dihapus Outlier.

Parameter	Jumlah	Mean	Std	Min	25%	50%	75%	Maks
Klorofil-a	45	0,2878	0,2077	0,013	0,127	0,227	0,401	0,748
Total Fosfor	45	27,44	8,8280	14	21	27	33	48
Secchi Disk	45	0,3484	0,0783	0,2	0,29	0,35	0,4	0,55

Penghapusan outlier memberikan dampak positif terhadap kualitas data yang digunakan dalam pemodelan. Data yang lebih bersih dan homogen dapat membantu proses Regresi *Random Forest* dalam mengidentifikasi pola secara lebih akurat, serta meningkatkan performa prediksi. Selain itu, penghapusan outlier dapat mengurangi resiko overfitting akibat nilai yang ekstrem yang tidak representatif terhadap populasi.

### Hasil Regresi Random Forest

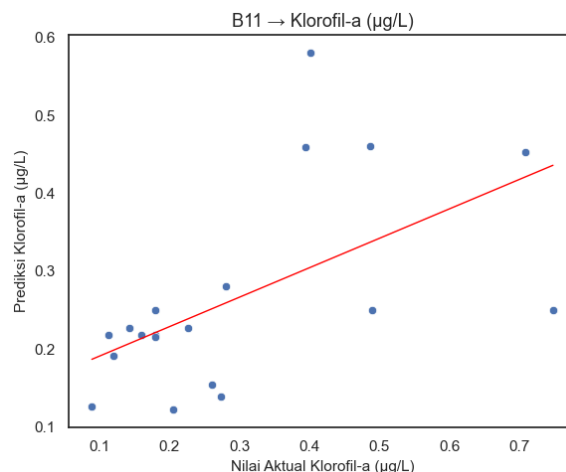
Performa model RRF menunjukkan peningkatan sesudah dilakukan *hyperparameter tuning* (Tabel 7). Hasil terbaik ditunjukkan oleh model yang telah di tuning. Dengan nilai  $R^2$  tertinggi pada B11 untuk parameter Klorofil-a dan Secchi Disk dengan nilai masing-masing adalah 0,2877 dan 0,3709. Sedangkan untuk Total Fosfor terdapat pada B8A dengan nilai 0,0277.

**Tabel 7** Hasil Regresi Random Forest

No	Band	Sebelum / Setelah <i>Hyperparamter Tuning</i>	Klorofil-a		Total Fosfor		Secchi Disk	
			$R^2$	RMSE	$R^2$	RMSE	$R^2$	RMSE
1	B2	Sebelum Tuning	0	0,2349	0	10,0695	0	0,0729
2		Setelah Tuning	0	0,1886	0	10,6891	0,1934	0,0642
3	B3	Sebelum Tuning	0	0,2252	0	11,6022	0,0246	0,0706
4		Setelah Tuning	0	0,2026	0	11,5507	0,0445	0,0699
5	B4	Sebelum Tuning	0	0,2584	0	13,6598	0,2345	0,0625
6		Setelah Tuning	0	0,2249	0	12,3671	0,2498	0,0619
7	B5	Sebelum Tuning	0,0602	0,1826	0	11,2191	0,1416	0,0662
8		Setelah Tuning	0,1088	0,1778	0	10,8156	0,2722	0,0610
9	B6	Sebelum Tuning	0	0,2236	0	11,3511	0	0,0784
10		Setelah Tuning	0	0,2225	0	10,7159	0	0,0744
11	B7	Sebelum Tuning	0	0,2027	0	11,0529	0	0,0733
12		Setelah Tuning	0	0,2105	0	10,2705	0	0,0722
13	B8	Sebelum Tuning	0	0,1903	0	10,7482	0,2218	0,0631
14		Setelah Tuning	0	0,1933	0	10,4398	0,3433	0,0579
15	B8A	Sebelum Tuning	0,1724	0,1713	0	10,3459	0,3618	0,0571
16		Setelah Tuning	0,0717	0,1814	0,0277	9,7197	0,2801	0,0606
17	B11	Sebelum Tuning	0,2710	0,1608	0	10,7321	0,3009	0,0598
18		Setelah Tuning	0,2877	0,1589	0	10,5986	0,3709	0,0567
19	B12	Sebelum Tuning	0,0322	0,1853	0	11,4452	0	0,0734
20		Setelah Tuning	0,2740	0,1604	0	10,8924	0,3181	0,0590

Sumber: Olah Data, 2025

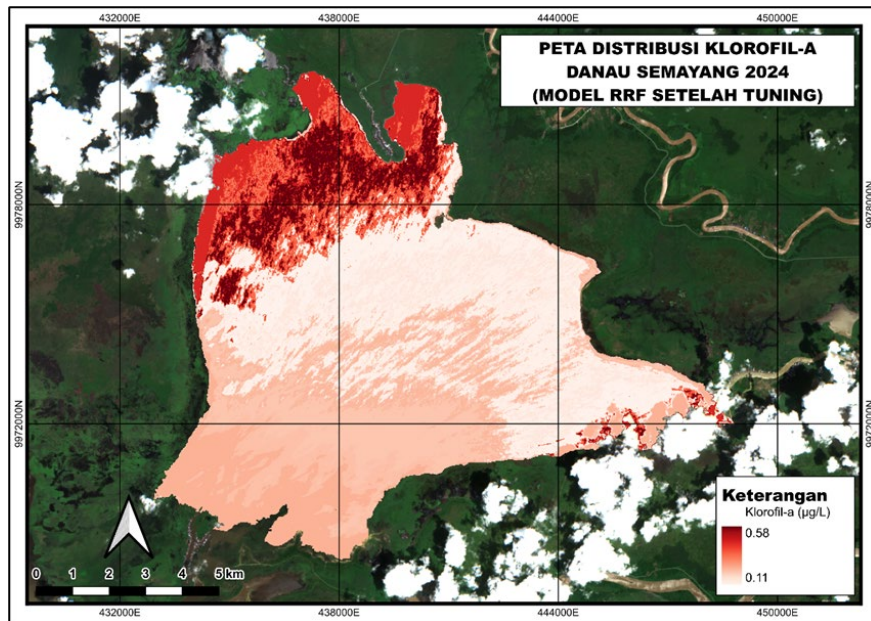
Scatter plot Klorofil-a (Gambar 6), menunjukkan adanya kecenderungan hubungan positif antara nilai aktual dan prediksi meskipun sebaran data masih cukup menyebar. Hal ini sesuai dengan  $R^2$  0,2877 yang menunjukkan bahwa B11 mampu merepresentasikan variasi Klorofil-a meskipun belum optimal. Dibandingkan dengan penelitian sebelumnya, hasil yang terdapat pada penelitian ini lebih rendah dari hasil didapatkan oleh Priadi & Sudaryatno (2016) yang memperoleh  $R^2$  sebesar 0,774 di Waduk Sutami menggunakan citra Landsat 8 OLI, sedangkan Rivani & Wicaksono (2018) mendapatkan nilai  $R^2$  sebesar 0,646 di Danau Maninjau, dan Musfiroh dkk (2020) mencapai 0,71 di Waduk Cirata menggunakan citra PlanetScope. Perbedaan ini disebabkan oleh perbedaan resolusi citra, nilai bit citra, karakteristik perairan, jumlah data dan metode yang digunakan. Namun, hasil studi ini masih lebih baik dari penelitian Amanatullah (2023) yang menggunakan citra Sentinel-2 MSI dan Regresi Random Forest namun hanya memperoleh nilai  $R^2$  sebesar 0,03 di Danau Singkarak. hal ini menunjukkan jika pendekatan dalam studi ini memiliki potensin, namun akurasi masih perlu ditingkatkan, dengan mengkombinasikan dengan multi-band atau indeks spektral.



**Gambar 6** Scatter plot Klorofil-a

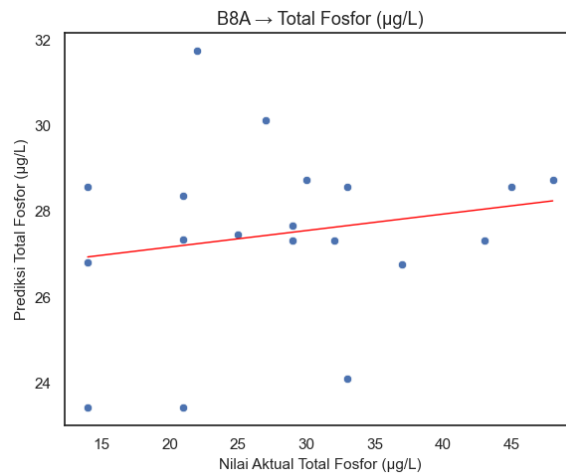
Peta distribusi Klorofil-a hasil prediksi model RRF (setelah tuning) menunjukkan sebaran spasial konsentrasi Klorofil-a di Danau Semayang pada tahun 2024 (Gambar 7). Nilai Klorofil-a berkisar antara 0,11–0,58 µg/L, dengan konsentrasi tertinggi terakumulasi di bagian utara dan timur laut dan cenderung menurun ke arah selatan dan tengah danau. Pola ini menunjukkan potensi eutrofikasi lokal yang lebih intens di wilayah tertentu

kemungkinan dipengaruhi oleh masukan nutrisi dari daratan atau aktivitas antropogenik di sekitar danau.



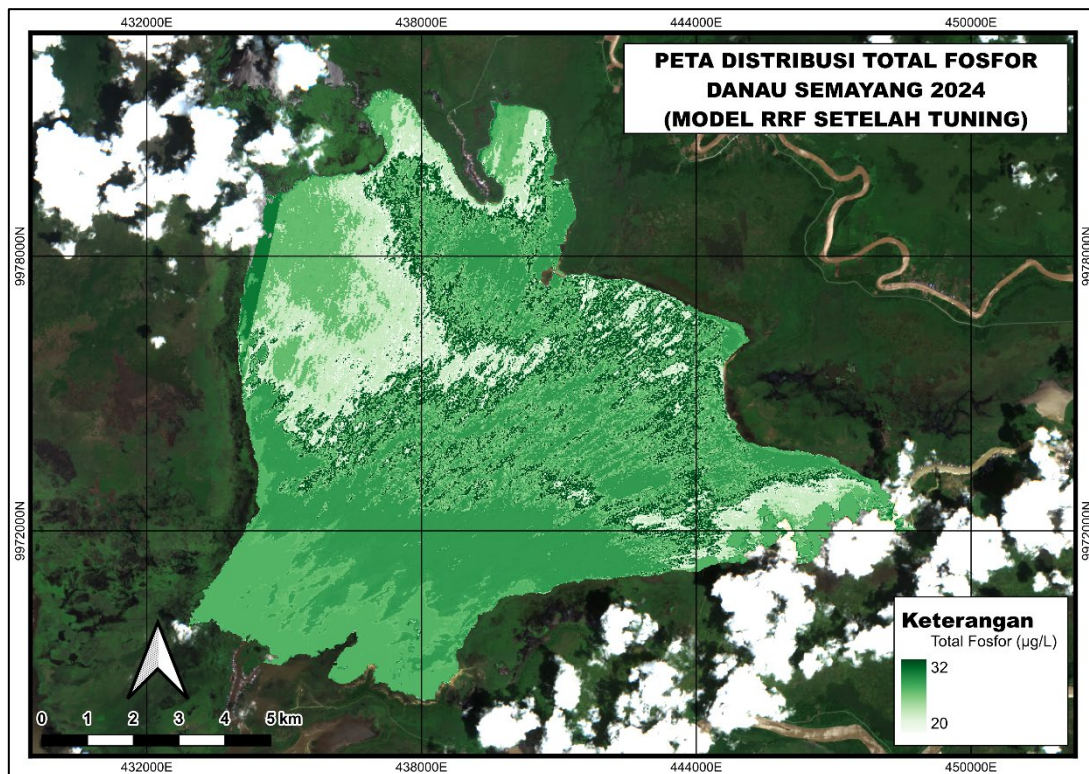
**Gambar 7** Peta Distribusi Klorofil-a

Parameter Total Fosfor, model menunjukkan kinerja rendah. Scatter Plot pada B8A (Gambar 8) menunjukkan pola yang sangat menyebar dengan slope garis regresi yang landai, dan didukung dengan nilai  $R^2$  yang rendah sebesar 0,0277. Hal tersebut mengindikasikan bahwa Total Fosfor sulit di prediksi hanya dari saluran tunggal, dan perlu dilakukan pendekatan multi-band atau data tambahan. Hasil ini lebih rendah dibandingkan studi Musfiroh dkk (2020) dan Rivani & Wicaksono (2018) yang masing-masing memperoleh  $R^2$  sebesar 0,65 dan 0,464 menggunakan citra resolusi tinggi seperti PlanetScope dan Landsat 8. Perbedaan ini kemungkinan disebabkan oleh keterbatasan jumlah dan distribusi sampel dalam studi ini, serta kompleksitas perairan yang tinggi. Hasil ini sejalan dengan temuan Amanatullah (2023) yang menggunakan Sentinel-2 dan Random Forest di Danau Singkarak dan memperoleh  $R^2$  sebesar 0, menguatkan bahwa prediksi Total Fosfor dari satu band reflektansi Sentinel-2 kurang memadai tanpa pendekatan multi-band atau data tambahan.



**Gambar 8** Scatter Plot Total Fosfor

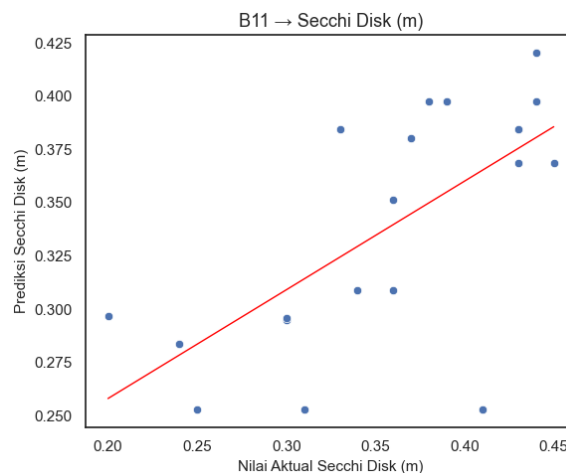
Peta Distribusi Total Fosfor tampak lebih merata dibandingkan parameter lainnya (Gambar 9), dengan kisaran 20–32  $\mu\text{g/L}$ . Nilai tertinggi terdistribusi di bagian utara dan tengah danau, yang mungkin dipengaruhi oleh masukan eksternal (seperti dari anak sungai) atau akumulasi di daerah yang lebih tenang. Namun, pola ini tidak sejelas distribusi Klorofil-a, mencerminkan prediksi model yang lebih lemah pada parameter ini. Keterbatasan performa model pada parameter ini membuat pola distribusi perlu diinterpretasi dengan hati-hati dan didukung oleh data lapangan.



**Gambar 9** Peta Distribusi Total Fosfor



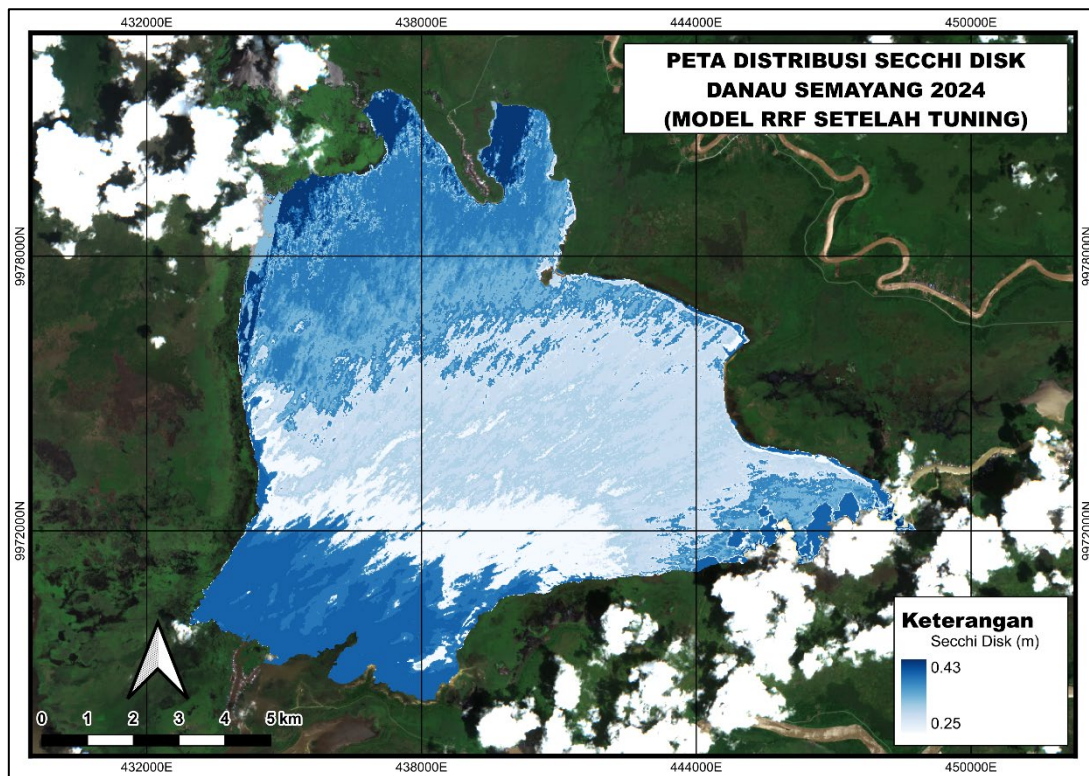
Prediksi terhadap Secchi Disk melalui B11 (Gambar 10) menunjukkan pola yang lebih baik dan linier. Titik-titik prediksi mengikuti arah garis regresi secara konsisten, dengan nilai  $R^2$  mencapai 0,3709. Hal tersebut menunjukkan bahwa kejernihan air lebih sensitif terhadap panjang gelombang SWIR dan B11 memberikan kontribusi terbesar dalam memodelkan variasi parameter ini. Priadi & Sudaryatno (2016b) melaporkan  $R^2$  sebesar 0,554, dan Musfiroh dkk (2020) sebesar 0,60, keduanya menggunakan citra resolusi tinggi. Meskipun masih di bawah capaian tersebut, hasil studi ini lebih baik dibandingkan Amanatullah (2023) ( $R^2 = 0,29$ ) dengan pendekatan serupa menggunakan Sentinel-2. Temuan ini menguatkan bahwa pemilihan band yang tepat, penghapusan outlier, dan pemodelan yang sesuai seperti RRF mampu meningkatkan akurasi prediksi, bahkan dalam kondisi perairan dinamis dan jumlah sampel terbatas. Meski belum melampaui capaian tertinggi Rivani & Wicaksono (2018) di Danau Maninjau, hasil ini menunjukkan potensi Sentinel-2 MSI dalam memetakan transparansi air dengan pendekatan yang tepat.



**Gambar 10** Scatter Plot Secchi Disk

Nilai transparansi air (Secchi Disk) berkisar 0,25–0,43 m, dengan kejernihan tertinggi terletak di bagian selatan dan tengah danau, dan kejernihan menurun ke arah utara (Gambar 11). Hal ini konsisten dengan distribusi Klorofil-a, di mana area dengan konsentrasi tinggi Klorofil-a cenderung memiliki kejernihan rendah akibat peningkatan kekeruhan. Memperkuat dugaan bahwa kekeruhan sebagian besar disebabkan oleh aktivitas fitoplankton atau partikel tersuspensi. SWIR (B11) terbukti sensitif dalam memodelkan parameter ini.

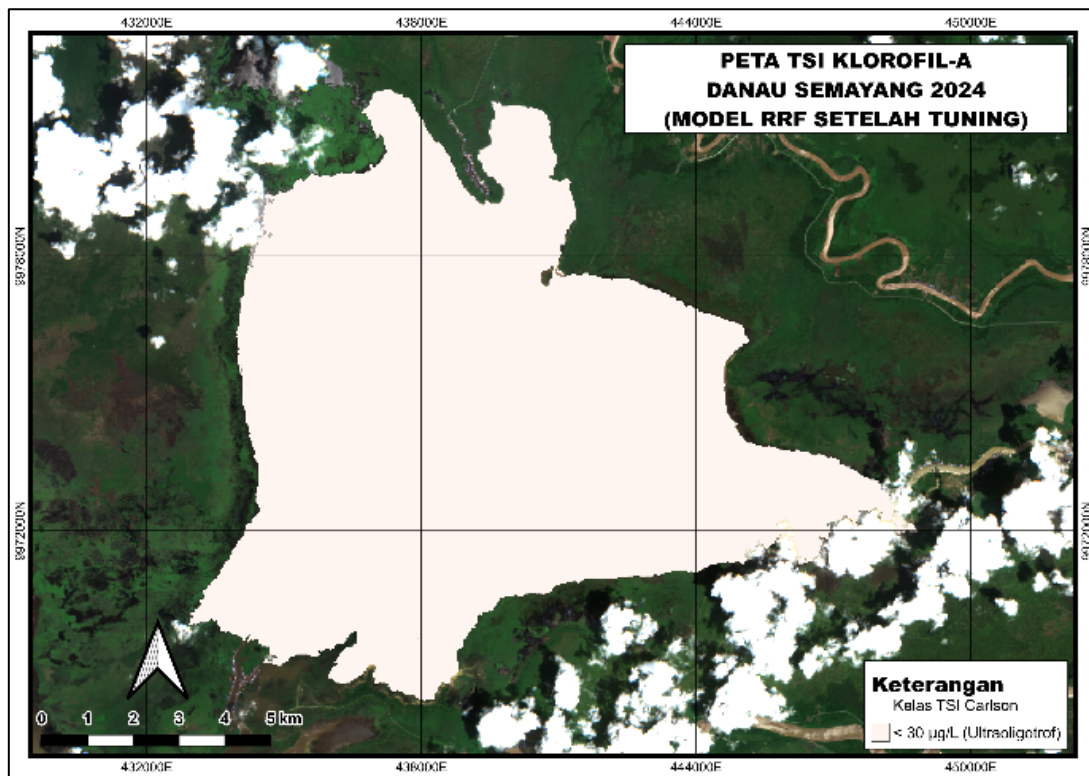




Gambar 11 Peta Distribusi Secchi Disk

### Pemetaan Status Trofik Carlson

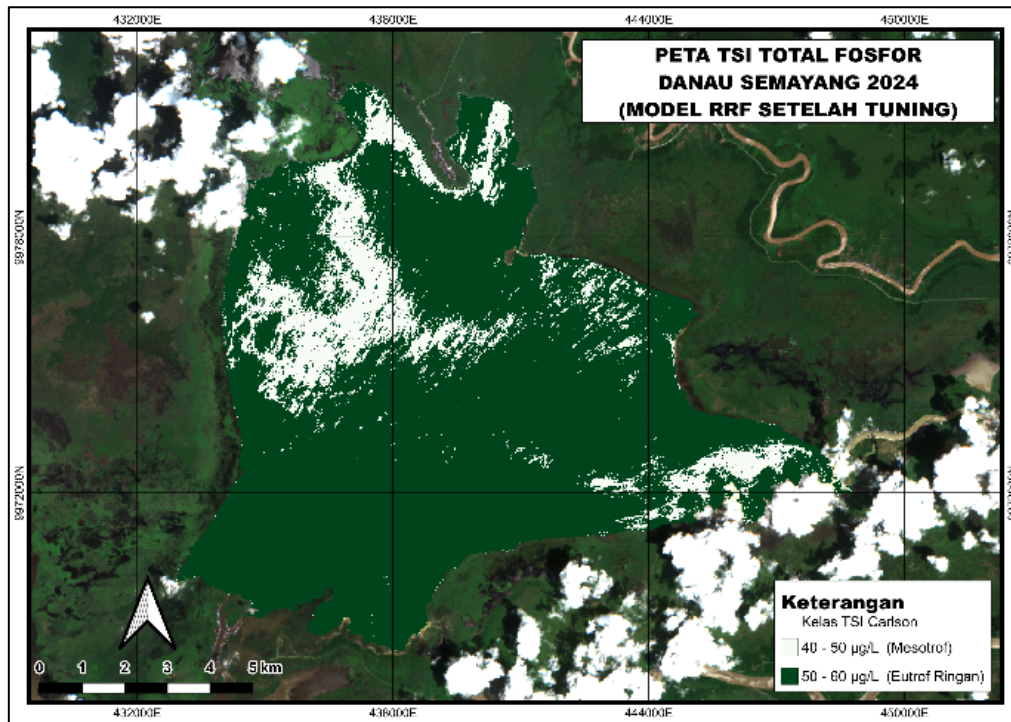
Pemetaan status trofik Danau Semayang dilakukan dengan mengonversi hasil estimasi parameter kualitas air (Klorofil-a, Total Fosfor, dan Secchi Disk) ke dalam indeks TSI Carlson. Nilai TSI dihitung menggunakan persamaan masing-masing parameter (Persamaan 1 - 3), sedangkan rata-rata status trofik diperoleh melalui Persamaan 4. Klasifikasi status trofik kemudian dilakukan berdasarkan kelas TSI Carlson (Tabel 2). Tahapan ini menghasilkan peta distribusi kelas status trofik yang merepresentasikan kondisi trofik danau secara spasial berdasarkan nilai TSI dari ketiga parameter utama.



**Gambar 12** Peta TSI Klorofil-a

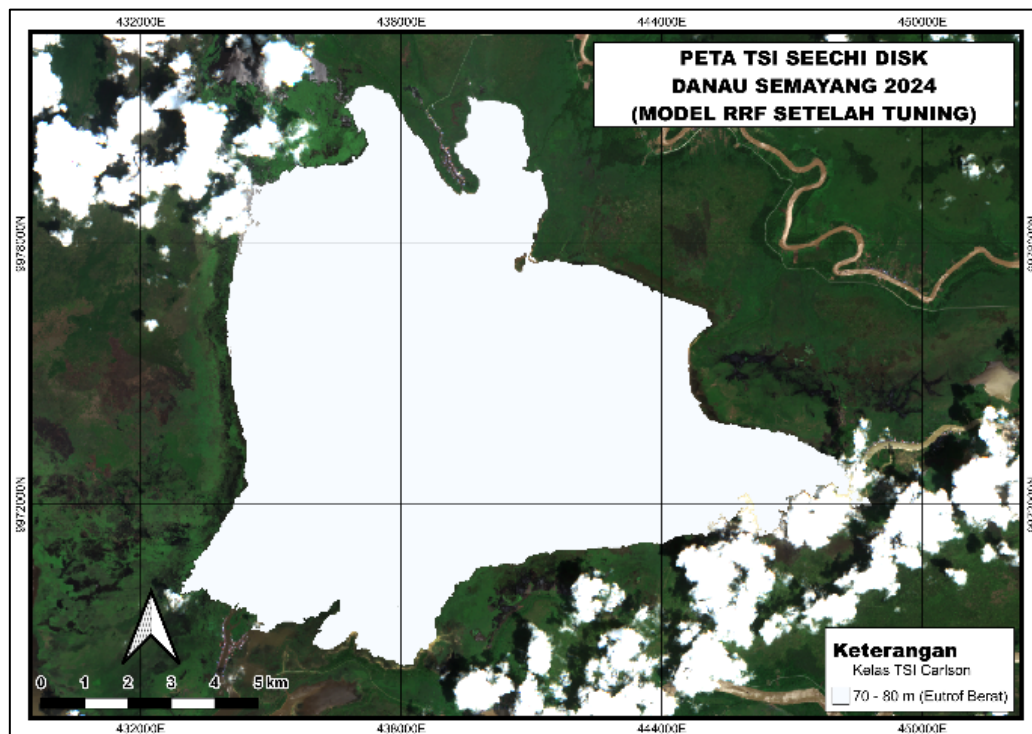
Seluruh wilayah Danau Semayang diklasifikasikan sebagai ultraoligotrof ( $TSI < 30$ ), yang menunjukkan tingkat produktivitas primer sangat rendah (Gambar 12). Hal ini sejalan dengan hasil prediksi Klorofil-a yang relatif kecil. Meskipun demikian, hasil ini perlu divalidasi lebih lanjut karena bisa jadi mencerminkan keterbatasan model dalam menangkap konsentrasi fitoplankton aktual.

Sebagian besar danau tergolong dalam kelas eutrof ringan ( $TSI 50\text{--}60$ ), sedangkan sebagian kecil masuk kategori mesotrof ( $TSI 40\text{--}50$ ) (Gambar 13). Kondisi ini menunjukkan bahwa meskipun klorofil rendah, ketersediaan fosfor cukup tinggi. Hal ini menjadi indikator bahwa danau memiliki potensi eutrofikasi apabila kondisi lingkungan mendukung pertumbuhan alga.



**Gambar 13** Peta TSI Total Fosfor

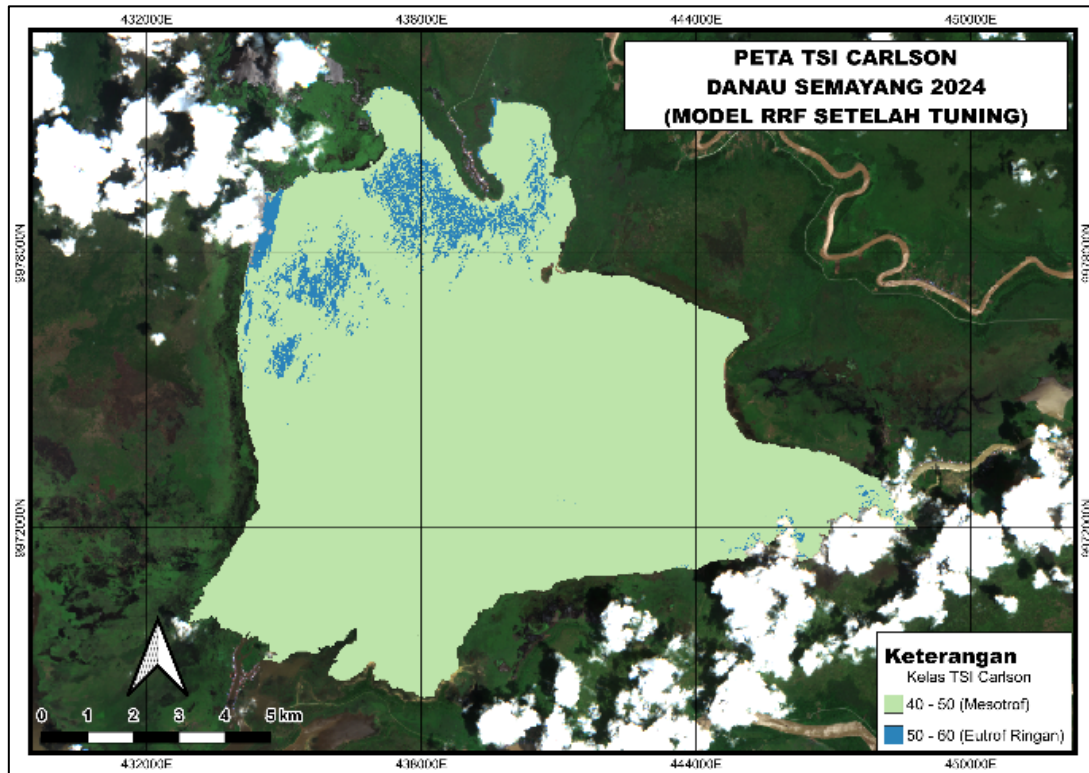
Seluruh area danau diklasifikasikan sebagai eutrof berat (TSI 70–80) berdasarkan indeks kejernihan. Hal ini menunjukkan tingkat kekeruhan tinggi yang tidak sebanding dengan nilai Klorofil-a. Kekeruhan kemungkinan besar disebabkan oleh partikel non-biologis seperti sedimen atau lumpur halus.



**Gambar 14** Peta TSI Secchi Disk



Sebagian besar wilayah Danau Semayang berada dalam kondisi mesotrofik, dengan beberapa area di utara dan timur laut menunjukkan kecenderungan eutrof ringan. Hal ini menggambarkan bahwa secara umum danau berada pada tingkat produktivitas sedang. Namun, adanya zona eutrofik mengindikasikan potensi penurunan kualitas air di area tertentu yang perlu mendapat perhatian.



Gambar 15 Peta Rata-rata TSI Carlson

### Uji Akurasi Hasil Konversi TSI Carlson

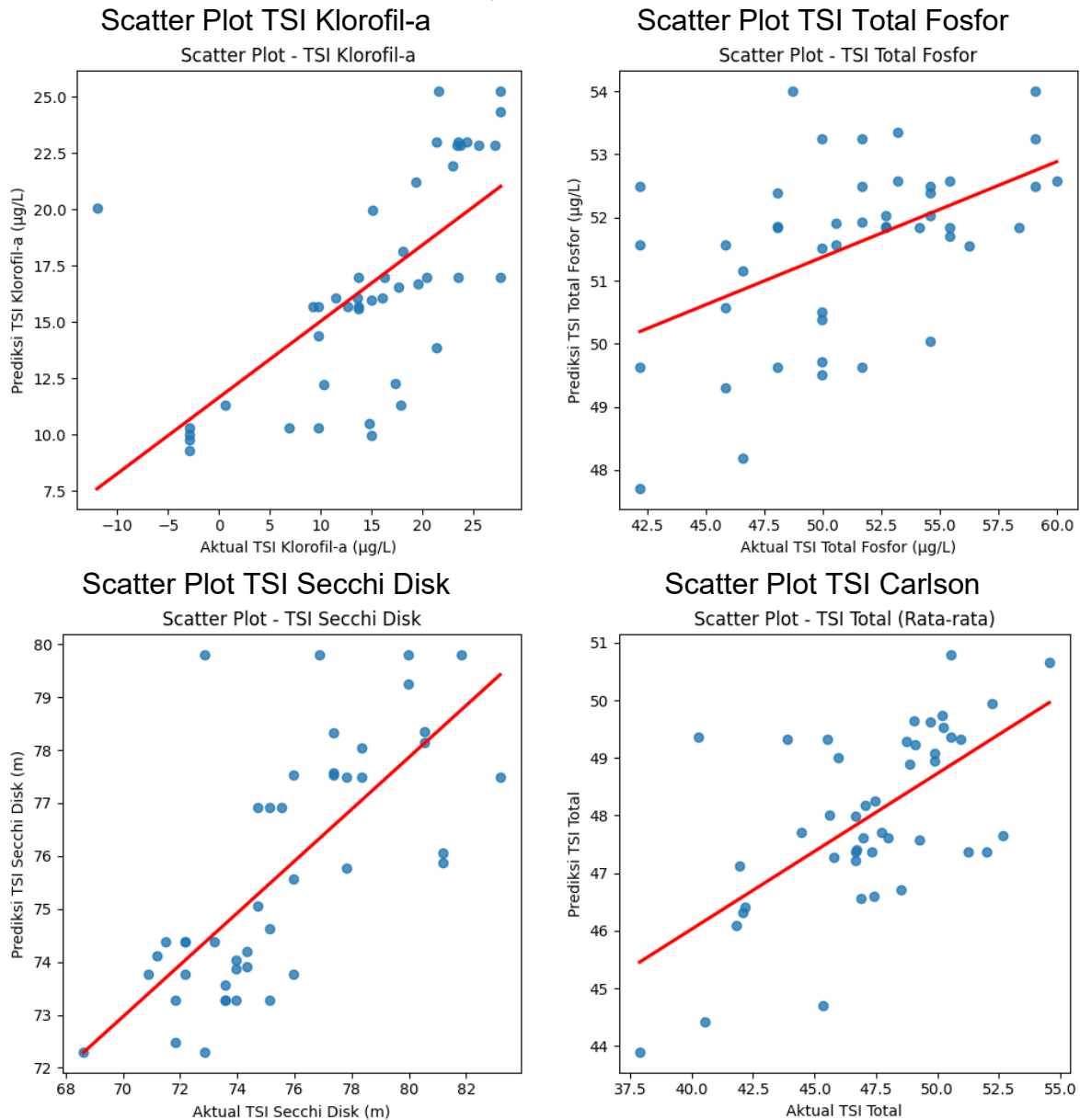
Hasil evaluasi konversi TSI Carlson menunjukkan variasi akurasi antar parameter (Tabel 8) yang digambarkan oleh scatter plot (Gambar 16). TSI Secchi Disk memiliki performa terbaik, terlihat dari sebaran titik yang mendekati garis regresi ideal ( $y = x$ ) dan nilai  $R^2$  tertinggi sebesar 0,5154, disertai RMSE paling rendah (2,32), yang menunjukkan prediksi model cukup presisi. Sementara itu, TSI Klorofil-a menunjukkan pola prediksi yang cenderung linier namun masih menyebar, sesuai dengan nilai  $R^2$  sedang (0,3793) dan RMSE tertinggi (7,30), menandakan masih adanya deviasi besar antara nilai aktual dan prediksi. TSI Total Fosfor menjadi parameter dengan akurasi terendah, ditunjukkan oleh sebaran titik yang tidak mengikuti pola jelas dan garis regresi yang datar, sejalan dengan  $R^2$  rendah (0,2045). Adapun TSI Total (rata-rata) memiliki akurasi sedang ( $R^2 = 0,3106$ ), mencerminkan bahwa penggabungan ketiga parameter memberikan hasil prediksi yang cukup representatif meski belum optimal. Scatter plot secara keseluruhan memperkuat

bahwa keberhasilan konversi TSI sangat tergantung pada sensitivitas masing-masing parameter terhadap prediktor yang digunakan.

**Tabel 8** Hasil Uji Akurasi Konversi TSI Carlson

No	Parameter	Evaluasi Matriks	
		$R^2$	RMSE
1	TSI Klorofil-a	0,3793	7,3041
2	TSI Total Fosfor	0,2045	4,1393
3	TSI Secchi Disk	0,5154	2,3217
4	TSI	0,3106	2.9318

Sumber: Olah Data, 2025



**Gambar 16** Scatter Plot Hasil Evaluasi Konversi TSI Carlson

Perbedaan tingkat akurasi antar parameter menunjukkan bahwa penginderaan jauh lebih sensitif dalam mendeteksi parameter optik seperti klorofil-a dan kejernihan perairan (secchi disk) dibandingkan dengan parameter kimiawi seperti total fosfor. Distribusi spasial Pemanfaatan *Random Forest* Dalam Pemetaan Status Trofik...  
Muhammad Dika Oktoberdinata, Pramaditya Wicaksono, Sudaryatno

klorofil-a dan secchi disk yang konsisten di bagian utara dan timur laut menunjukkan adanya masukan nutrisi dari anak sungai yang mengarah ke danau dan aktivitas antropogenik di sekitar danau. Temuan ini sejalan dengan karakteristik danau paparan banjir yang memiliki sifat dinamis, dan kondisi danau sangat dipengaruhi oleh masukan dari sungai utama atau anak sungai yang bermuara ke Danau Semayang.

Kondisi mesotrofik dengan kecenderungan eutrof ringan perlu mendapat perhatian, karena berpotensi berkembang menjadi eutrof berat yang dapat menurunkan kualitas perairan dan mengancam fungsi ekosistem Danau Semayang. Oleh karena itu, pengembangan pendekatan multi-band, pemanfaatan indeks spektral, serta integrasi data hidrologi dan meteorologi menjadi penting untuk meningkatkan performa model. Selain itu, eksplorasi algoritma pembelajaran mesin lain seperti BPNN atau XGBoost dapat memberikan alternatif yang lebih akurat dalam pemodelan status trofik di ekosistem perairan yang kompleks.

## Kesimpulan

Penelitian ini berhasil mengimplementasikan algoritma Regresi *Random Forest* untuk memetakan status trofik Danau Semayang menggunakan citra Sentinel-2 MSI. Model terbaik diperoleh setelah proses *hyperparameter tuning*, dengan Band 11 (SWIR) menunjukkan performa tertinggi dalam memodelkan parameter Klorofil-a ( $R^2 = 0,2877$ ; RMSE = 0,1589) dan kejernihan (Secchi Disk) ( $R^2 = 0,3709$ ; RMSE = 0,0567). Sementara itu, prediksi Total Fosfor masih rendah akurasi ( $R^2 = 0,0277$ ; RMSE = 9,7197), menunjukkan perlunya pendekatan *multi-band* atau penambahan variabel lain. Hasil konversi ke *Trophic State Index* (TSI) Carlson menunjukkan bahwa mayoritas wilayah Danau Semayang tergolong mesotrofik (TSI rata-rata), dengan kecenderungan eutrof ringan di bagian utara dan timur laut. Temuan ini mengindikasikan potensi penurunan kualitas air lokal. Metode berbasis penginderaan jauh ini terbukti efisien untuk pemantauan kualitas air di danau paparan banjir dengan kondisi yang dinamis.

## Daftar Pustaka

- Amanatullah, D. A. (2023). *Pemetaan Kondisi Status Trofik Danau Singkarat Provinsi Sumatera Barat Menggunakan Citra Sentinel-2 MSI*. Universitas Gadjah Mada.
- Ariyanto, E., Karim, M. A., & Firmansyah, A. (2014). Biobriket Enceng Gondok (*Eichhornia Crassipes*) Sebagai Bahan Bakar Energi Terbarukan. *Reaktor*, 15(1), 59. <https://doi.org/10.14710/reaktor.15.1.59-63>
- Asdak, C. (2007). *Hidrologi dan Pengelolaan Daerah Aliran Sungai*. Gadjah Mada University Press.
- Carlson, R. E. (1977). A trophic state index for lakes. *Limnology and Oceanography*, 22(2), 361–369. <https://doi.org/10.4319/lo.1977.22.2.0361>
- Pemanfaatan *Random Forest* Dalam Pemetaan Status Trofik...
- Muhammad Dika Oktoberdinata, Pramaditya Wicaksono, Sudaryatno

- Darajati, W., Pratiwi, S., Herwinda, E., Radiansyah, A. D., Nalang, V. S., Nooryanto, B., Rahajoe, J. S., Ubaidillah, R., Maryanto, I., Kurniawan, R., Prasetyo, T. A., Rahim, A., Jefferson, J., & Hakim, F. (2016). Indonesia Biodiversity Strategi and Action Plan 2015-2020. In *Kementrian Perencanaan Pembangunan Nasional/ BAPPENAS*. Kementerian Perencanaan Pembangunan Nasional.
- ESA. (2015). *Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services*. ESA Communications Production. [www.esa.int](http://www.esa.int)
- Fu, P., Li, X., Zhang, J., Ma, C., Wang, Y., & Meng, F. (2024). Remote Sensing Inversion on Heavy Metal Content in Salinized Soil of Yellow River Delta Based on Random Forest Regression—a Case Study of Gudao Town. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-62087-y>
- Islam, K. I., Elias, E., Carroll, K. C., & Brown, C. (2023). Exploring Random Forest Machine Learning and Remote Sensing Data for Streamflow Prediction: An Alternative Approach to a Process-Based Hydrologic Modeling in a Snowmelt-Driven Watershed. *Remote Sensing*, 15(16). <https://doi.org/10.3390/rs15163999>
- Klippel, G., Macêdo, R. L., & Branco, C. W. C. (2020). Comparison of different trophic state indices applied to tropical reservoirs. *Lakes and Reservoirs: Science, Policy and Management for Sustainable Use*, 25(2), 214–229. <https://doi.org/10.1111/lre.12320>
- Kutser, T., Hedley, J., Giardino, C., Roelfsema, C., & Brando, V. E. (2020). Remote sensing of shallow waters – A 50-year retrospective and future directions. *Remote Sensing of Environment*, 240(December 2019), 111619. <https://doi.org/10.1016/j.rse.2019.111619>
- Liu, N., Hu, Y., & Ai, X. (2019). Research on Power Load Forecasting Based on Random Forest Regression. *IOP Conference Series: Earth and Environmental Science*, 252(3). <https://doi.org/10.1088/1755-1315/252/3/032171>
- Lutfhi, W. (2020). Data Danau Diperbarui, Kini Ada 5.807 Danau yang Tersebar di Indonesia. In *goodnewsfromindonesia.id*. <https://www.goodnewsfromindonesia.id/2021/09/02/data-danau-diperbarui-kini-ada-5807-danau-yang-tersebar-di-indonesia>
- Mahmud, S. L., & Achmad, N. (2020). Analisis Dinamik Model Pendangkalan Danau Limboto Dengan Pembersihan Eceng Gondok Dan Pengerukan Endapan. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 14(4), 597–608. <https://doi.org/10.30598/barekengvol14iss4pp597-608>
- Meng, D., Mao, J., Li, W., Zhu, S., & Gao, H. (2023). An Integrated Framework for Remote Sensing Assessment of the Trophic State of Large Lakes. *Remote Sensing*, 15(17). <https://doi.org/10.3390/rs15174238>
- Mislan, Hendra, M., Lariman, Trimurti, S., Anwar, Y., Rahman, N., Iskandar, A., Pratama, A., & Aprianti, D. A. (2022). *Laporan Studi Keanekaragaman Hayati di Danau Kaskade Mahakam Tahun 2022*.
- Musfiroh, Z., Ridwansyah, I., & Murti, S. H. (2020). The application of PlanetScope imagery to map out the trophic state of Cirata Reservoir, West Java. *IOP Conference Series: Earth and Environmental Science*, 535(1). <https://doi.org/10.1088/1755-1315/535/1/012032>
- Panitia Kerja Penyusunan Rencana Pengelolaan Danau Kaskade Mahakam. (2019). *Rencana Pengelolaan Danau Kaskade Mahakam*. Kementerian Lingkungan Hidup dan Kehutanan.
- Priadi, T., & Sudaryatno. (2016). *Identifikasi Status Trofik Perairan Menggunakan Data Penginderaan Jauh Citra Landsat-8 OLI (Operationa Land Imager) di Waduk Sutami (Karangkates) Jawa Timur*. 1–23.
- Rivani, A., & Wicaksono, P. (2018). Water Trophic Status Mapping of Tecto-Volcanic Maninjau Lake during Algae Bloom using Landsat 8 OLI Satellite Imagery. *ICARES 2018 - Proceedings of the 2018 IEEE International Conference on Aerospace*



*Electronics and Remote Sensing Technology*, 127–133.  
<https://doi.org/10.1109/ICARES.2018.8547055>

Zhou, Z., Qiu, C., & Zhang, Y. (2023). A comparative analysis of linear regression, neural networks and random forest regression for predicting air ozone employing soft sensor models. *Scientific Reports*, 13(1), 1–23. <https://doi.org/10.1038/s41598-023-49899-0>