



Analisis Prediksi Sebaran Zona Potensial Penangkapan Ikan (ZPPI) Menggunakan Algoritma K-Nearest Neighbor di Perairan Gunung Kidul

Sigit Nur Dian Syah¹, Ozzi Surya²

^{1,2} Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta
e-mail: sigitnurdiansyah18@gmail.com, ozzisuria@mercubuana-yogya.ac.id

Abstract

The southern waters of Gunung Kidul Regency possess abundant capture fisheries resources; however, traditional fishers still face difficulties in determining effective fishing locations, which are often selected speculatively. This study aims to design a computational spatial prediction system for Potential Fishing Zones using the *K-Nearest Neighbor* (KNN) algorithm. The primary dataset was extracted from daily NOAA satellite imagery in NetCDF format covering the period from 15 to 31 October 2025 and representing five major fish auction site clusters. Data preprocessing included spatial filtering, removal of data affected by cloud cover, separation of predictor features, and metric transformation using *Min-Max Scaling*. Actual data labeling was performed based on the *Standard Oceanographic Threshold*, in which an area was classified as potentially productive when the Sea Surface Temperature (SST) was $\geq 28.0^{\circ}\text{C}$ and the chlorophyll-a concentration exceeded 0.22 mg/m^3 . Computational experiments were conducted using a *10-Fold Cross Validation* scenario by comparing *Euclidean Distance* and *Manhattan Distance* metrics across odd K values of 3, 5, 7, 9, and 11. The results demonstrate that the KNN model achieved its best spatial mapping performance using K=3 with *Euclidean Distance*, producing a final accuracy of 97% and an *F1-Score* of 0.97. The prediction outputs were then integrated into an interactive *Web-GIS* interface using the Folium library in Python 3.12. The system visually distinguishes potential and non-potential fishing areas, indicating that KNN can serve as an effective, precise, and data-driven decision-support tool for coastal fishers in identifying promising fishing locations.

Keywords: *Cross-Validation*; *K-Nearest Neighbor*; Chlorophyll-a; Sea Surface Temperature; *Web-GIS*.

Abstrak

Wilayah perairan selatan Kabupaten Gunung Kidul memiliki potensi sumber daya perikanan tangkap yang melimpah, namun pemanfaatannya oleh nelayan tradisional masih menghadapi kendala dalam menentukan lokasi penangkapan ikan yang efektif dan tidak bersifat spekulatif. Penelitian ini bertujuan untuk merancang sistem prediksi spasial Zona Potensial Penangkapan Ikan (ZPPI) secara komputasional menggunakan algoritma *K-Nearest Neighbor* (KNN). Dataset utama diperoleh dari citra satelit NOAA berformat NetCDF secara harian selama periode 15–31 Oktober 2025 yang mencakup lima kluster Tempat Pelelangan Ikan (TPI) utama. Tahapan prapemrosesan meliputi penyaringan spasial, pembersihan data akibat tutupan awan, pemisahan fitur prediktor, dan transformasi nilai menggunakan metode *Min-Max Scaling*. Pelabelan data aktual dilakukan berdasarkan *Standard Oceanographic Threshold*, yaitu area dikategorikan potensial apabila Suhu Permukaan Laut (SST) $\geq 28,0^{\circ}\text{C}$ dan konsentrasi Klorofil-a $> 0,22 \text{ mg/m}^3$. Eksperimen dilakukan menggunakan skenario *10-Fold Cross Validation* dengan membandingkan metrik *Euclidean Distance* dan *Manhattan Distance* pada nilai K=3, 5, 7, 9, dan 11. Hasil penelitian menunjukkan bahwa konfigurasi K=3 dengan *Euclidean Distance* menghasilkan performa terbaik, dengan akurasi sebesar 97% dan *F1-Score* sebesar 0,97. Hasil prediksi selanjutnya diintegrasikan ke dalam antarmuka *Web-GIS* interaktif menggunakan pustaka Folium pada Python 3.12. Sistem menampilkan area potensial dan non-potensial secara visual, sehingga dapat mendukung nelayan dalam menentukan lokasi penangkapan ikan secara lebih efektif, presisi, dan berbasis data.

Kata Kunci: *Cross-Validation*; *K-Nearest Neighbor*; Klorofil-A; Suhu Permukaan Laut; *Web-GIS*.

1. Pendahuluan

Wilayah perairan laut Indonesia merupakan salah satu koridor perikanan tangkap paling dinamis di kawasan ekuator, di mana kelestarian serta eksploitasi perikanan menjadi tulang punggung ekonomi masyarakat pesisir [3]. Salah satu area penangkapan ikan pelagis yang sangat strategis terletak di perairan Samudra Hindia yang berbatasan langsung dengan pesisir Kabupaten Gunung Kidul. Secara oseanografi, perairan ini dipengaruhi oleh fenomena *upwelling* yang membawa massa air dingin kaya nutrisi dari dasar laut ke permukaan sehingga memicu tingginya produktivitas primer daerah tersebut [1]. Meskipun memiliki potensi melimpah, mayoritas nelayan lokal masih beroperasi menggunakan insting tradisional yang spekulatif. Hal ini secara langsung memicu terjadinya inefisiensi jam melaut, pembengkakan biaya operasional bahan bakar minyak pelayaran, serta penurunan laju hasil tangkapan harian secara signifikan akibat ketidakpastian lokasi pergerakan gerombolan ikan.

Di era transformasi digital saat ini, teknologi penginderaan jauh menawarkan solusi mutakhir untuk memantau dinamika kelautan secara seketika dan cakupan yang luas. Dua parameter satelit utama yang telah divalidasi memiliki korelasi kuat terhadap konsentrasi kawanan ikan pelagis adalah Suhu Permukaan Laut (SST) [2] dan distribusi spasial konsentrasi Klorofil-a [3]. Untuk mentransformasi besaran data spasial berformat *NetCDF* yang

Analisis Prediksi Sebaran Zona Potensial Penangkapan Ikan (ZPPI) Menggunakan Algoritma K-Nearest Neighbor di Perairan Gunung Kidul

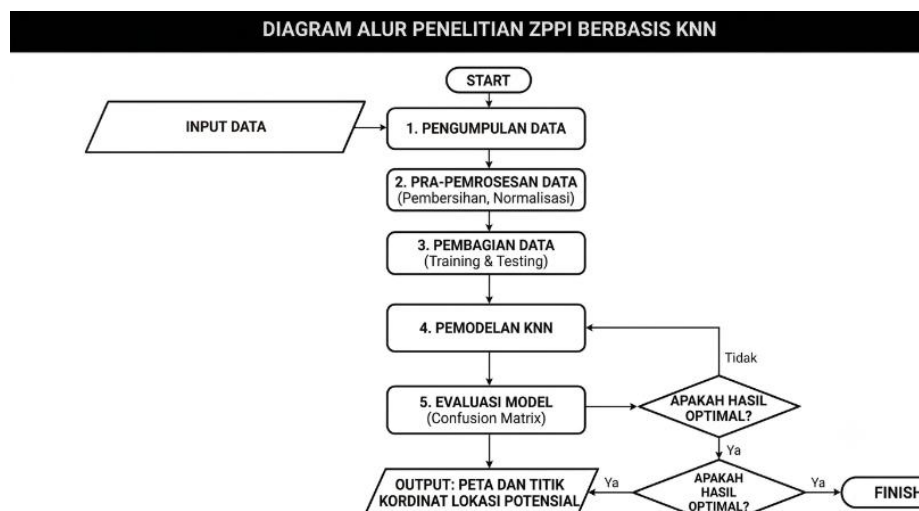
sangat kompleks tersebut menjadi instrumen keputusan penangkapan yang akurat secara otomatis, diperlukan intervensi dari sains komputasi melalui pendekatan *machine learning*. Dalam ranah klasifikasi data spasial multivariabel, algoritma *K-Nearest Neighbor* (KNN) terbukti secara saintifik andal untuk membangun model prediksi kawasan laut potensial karena sifat regresi non-parametriknya yang tangguh [4], sangat efisien untuk pemetaan data terpusat lokal, serta stabil terhadap sebaran anomali cuaca di permukaan laut [5].

Landasan fundamental untuk pelabelan daerah tangkapan ideal didasarkan pada karakteristik ambang batas kesuburan perairan tropis Indonesia, di mana rentang batas SST dan distribusi konsentrasi Klorofil-a menjadi indikator utama kehadiran fitoplankton [2]. Pemrosesan komputasi dalam penelitian ini difokuskan pada ekstraksi data harian satelit NOAA lokal secara riil yang direkam khusus pada rentang waktu 15 hingga 31 Oktober 2025 di sepanjang pesisir Gunung Kidul. Berdasarkan latar belakang tersebut, penelitian ini merumuskan masalah mengenai mekanisme prapemrosesan data satelit, komparasi pengukuran metrik jarak matematis KNN, optimalisasi *hyperparameter* K, serta bagaimana merancang visualisasi peta navigasi interaktif. Adapun tujuan utama dari riset ini adalah mengukur dan memvalidasi tingkat akurasi model KNN melalui skenario *10-Fold Cross Validation* serta menyediakan luaran aplikasi peta digital interaktif berbasis *Web-GIS* menggunakan pustaka *Folium* sebagai instrumen navigasi taktis bagi komunitas nelayan.

2. Metode Penelitian

Bagian ini menguraikan tahapan metodologi yang digunakan dalam penelitian untuk mengekstraksi dan mengklasifikasikan wilayah perairan menjadi Zona Potensi Penangkapan Ikan (ZPPI). Pendekatan yang digunakan adalah penelitian kuantitatif komputasi eksperimental yang bertumpu pada analisis data geospasial. Secara garis besar, pembahasan pada bagian ini mencakup dua pilar utama, yaitu: desain arsitektur dan alur kerja pemrosesan sistem beserta karakteristik ekstraksi dataset satelit NOAA (2.1), serta perancangan skenario pengujian model algoritma *K-Nearest Neighbor* (2.2).

2.1. Prosedur dan Alur Penelitian



Gambar 1. Diagram Alur Penelitian

Sebagaimana diilustrasikan pada Gambar 1, penelitian ini dilaksanakan melalui empat tahapan komputasi utama yang saling berurutan, yaitu pengumpulan data, pra-pemrosesan data, pembagian data, dan pemodelan K-Nearest Neighbor (KNN). Setiap tahapan dijabarkan secara rinci pada sub-bagian berikut, sedangkan tahap evaluasi model dibahas lebih lanjut pada bagian Hasil dan Pembahasan (Bagian 3).

2.1.1. Pengumpulan Data

Tahap pertama adalah pengumpulan data, yaitu proses integrasi data sekunder penginderaan jauh berupa citra satelit NOAA berformat NetCDF dengan data lapangan (ground truth) pada rentang waktu 15 hingga 31 Oktober 2025 di perairan selatan Kabupaten Gunung Kidul. Citra satelit harian ini memuat rekaman parameter Suhu Permukaan Laut (SST) dan konsentrasi Klorofil-a (CHL) yang menjadi variabel prediktor utama, sedangkan data lapangan digunakan sebagai label aktual status perairan berdasarkan kriteria Standard Oceanographic Threshold. Sebagai ilustrasi konkret dari data citra satelit yang dikumpulkan, kedua parameter oseanografi diakses dan diunduh melalui portal ERDDAP NOAA CoastWatch dalam format tabular (CSV) untuk keperluan inspeksi awal sebelum diolah lebih lanjut. Dua produk data yang digunakan yaitu produk suhu permukaan laut `jplMURSST41mday` dan produk konsentrasi klorofil-a `noaacwN20VIIRSchlaDaily`. Setiap berkas CSV memuat kolom time (waktu perekaman dalam format UTC), latitude dan longitude (koordinat titik piksel dalam satuan

derajat), serta nilai parameter oseanografi yang bersangkutan, yaitu sst dalam satuan derajat Celcius pada Tabel 1 dan chlor_a dalam satuan mg/m³ pada Tabel 2.

Tabel 1. Cuplikan Data Citra Suhu Permukaan Laut (SST) Produk jplMURSST41mday

No.	time (UTC)	latitude (°)	longitude (°)	sst (°C)	nobs
1	2025-10-16T00:00:00Z	-8.35	110.40	28.489044	31
2	2025-10-16T00:00:00Z	-8.35	110.41	28.478150	31
3	2025-10-16T00:00:00Z	-8.35	110.42	28.463715	31
4	2025-10-16T00:00:00Z	-8.35	110.43	28.446533	31
5	2025-10-16T00:00:00Z	-8.35	110.44	28.428833	31

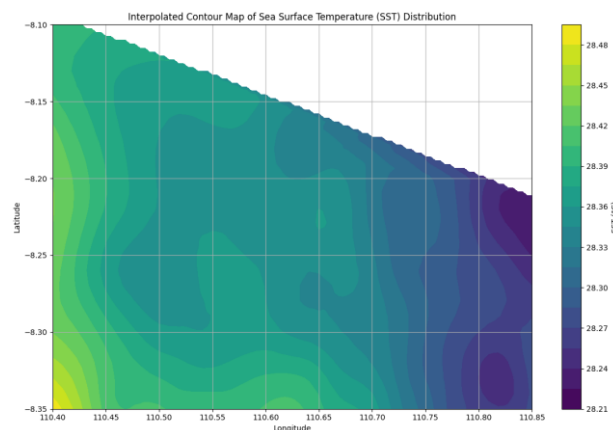
Tabel 2. Cuplikan Data Citra Konsentrasi Klorofil-a Produk noaacwN20VIIRSchlaDaily

No.	time (UTC)	altitude (m)	latitude (°)	longitude (°)	chlor_a (mg/m ³)
1	2025-10-16T12:00:00Z	0.0	-8.08125	110.41875	2.276286
2	2025-10-16T12:00:00Z	0.0	-8.08125	110.45625	NaN
3	2025-10-16T12:00:00Z	0.0	-8.08125	110.49375	NaN
4	2025-10-16T12:00:00Z	0.0	-8.08125	110.53125	NaN
5	2025-10-16T12:00:00Z	0.0	-8.08125	110.56875	NaN

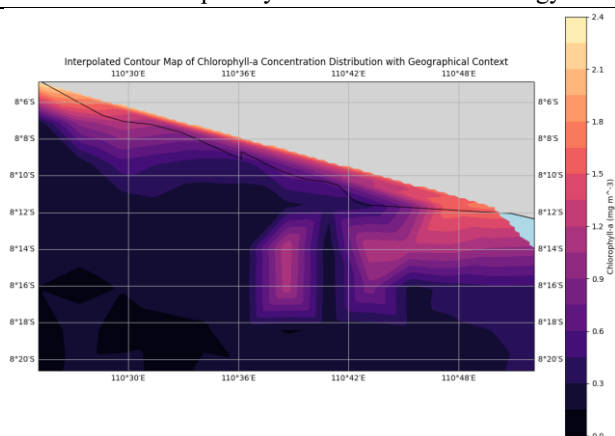
Sebagaimana tampak pada kedua tabel, tidak seluruh piksel memiliki nilai valid. Sel bertanda NaN pada Tabel 2 menunjukkan piksel yang gagal terekam akibat tutupan awan atau berada di luar jangkauan sensor pada waktu perekaman tersebut, sehingga baris data semacam ini perlu disaring pada tahap pra-pemrosesan sebelum digunakan untuk membentuk dataset tabular final (Latitude, Longitude, SST, Klorofil-a, dan label status perairan) yang diuraikan pada sub-bagian berikutnya.

2.1.2. Pra-Pemrosesan Data

Tahap kedua adalah pra-pemrosesan data, yang bertujuan menyiapkan data mentah agar layak diproses oleh algoritma klasifikasi jarak matematis. Tahapan ini meliputi penyaringan spasial untuk membatasi cakupan area kajian, pembersihan piksel citra yang terkontaminasi tutupan awan, pemisahan fitur prediktor (SST dan Klorofil-a) dari struktur data NetCDF multidimensi, serta transformasi nilai metrik menggunakan metode Min-Max Scaling agar seluruh variabel berada pada rentang skala yang setara. Contoh visualisasi citra hasil ekstraksi satelit NOAA yang telah melalui tahap interpolasi spasial ditunjukkan pada Gambar 2 untuk parameter Suhu Permukaan Laut (SST) dan Gambar 3 untuk parameter konsentrasi Klorofil-a.



Gambar 2. Peta Kontur Sebaran Suhu Permukaan Laut (SST) Perairan Gunung Kidul, 15–31 Oktober 2025



Gambar 3. Peta Kontur Sebaran Konsentrasi Klorofil-a Perairan Gunung Kidul, 15–31 Oktober 2025

Berdasarkan Gambar 2, sebaran SST menunjukkan gradasi suhu yang menurun secara gradual dari wilayah perairan bagian barat menuju bagian timur, dengan rentang nilai sekitar 28,21–28,49°C. Pola ini konsisten dengan indikasi fenomena upwelling lokal yang membawa massa air lebih dingin ke permukaan pada bagian timur cakupan kajian. Sementara itu, Gambar 3 memperlihatkan sebaran konsentrasi Klorofil-a yang bervariasi antara 0 hingga lebih dari 2,0 mg/m³, dengan beberapa titik konsentrasi tinggi (hotspot) yang muncul di dekat garis pantai. Titik-titik hotspot inilah yang menjadi salah satu indikator utama penentuan area dengan potensi kesuburan perairan yang tinggi.

Dataset lingkungan perairan dalam penelitian ini diekstraksi dari citra satelit NOAA harian yang direkam pada rentang periode 15 hingga 31 Oktober 2025. Data tersebut diunduh secara resmi melalui portal NOAA CoastWatch dengan fokus area pada ruang perairan selatan Kabupaten Gunung Kidul. Proses ekstraksi spasial dilakukan terhadap berkas data spasial berformat NetCDF atau berkode ekstensi dot nc menggunakan bantuan operasi pustaka Xarray. Dua parameter utama yang diturunkan dari pembacaan sensor satelit ini adalah nilai Suhu Permukaan Laut (SST) dan konsentrasi Klorofil-a (CHL). Untuk komponen data validasi lapangan atau data keadaan riil (ground truth), eksperimen ini menerapkan pendekatan kriteria ambang batas baku kesuburan laut (Standard Oceanographic Threshold). Metode ini menetapkan label potensi wilayah berdasarkan parameter bio-fisik perairan secara objektif [3]. Hasil ekstraksi spasial tersebut kemudian ditransformasikan ke dalam struktur dataset tabular yang memuat kolom koordinat garis lintang (Latitude), koordinat garis bujur (Longitude), nilai parameter SST dalam satuan derajat Celcius, kadar CHL dalam satuan miligram per meter kubik, serta label target status perairan. Variabel target atau kelas biner diberi nilai angka satu untuk merepresentasikan Zona Potensial dan diberi nilai angka nol untuk mengidentifikasi Zona Tidak Potensial.

2.1.3. Pembagian Data

Tahap ketiga adalah pembagian data, yaitu proses pembagian dataset biner secara acak menggunakan fungsi `train_test_split` dari pustaka Scikit-Learn dengan nilai `random state` tetap agar hasil pembagian bersifat konsisten pada setiap pengulangan eksperimen. Melalui fungsi tersebut, dataset dipisahkan menjadi data latih (training set) yang digunakan untuk melatih pemahaman model dan data uji (testing set) yang digunakan untuk mengevaluasi kemampuan generalisasi model, dengan rincian proporsi rasio pembagian data dijabarkan lebih lanjut pada skenario pengujian Bagian 2.2.

2.1.4. Pemodelan K-Nearest Neighbor (KNN)

Tahap keempat adalah pemodelan K-Nearest Neighbor (KNN), yaitu proses klasifikasi terhadap objek titik perairan baru berdasarkan kedekatan karakteristik lingkungannya terhadap data historis. Luaran dari tahap ini kemudian diuji tingkat keandalannya melalui instrumen Confusion Matrix, di mana rincian metrik evaluasi dan hasil optimasi hyperparameter K dijabarkan lebih lanjut pada Bagian 3 (Hasil dan Pembahasan).

Mengingat rentang angka nilai operasional SST yang berada di skala puluhan derajat memiliki ketimpangan variansi yang sangat besar jika dibandingkan dengan nilai CHL yang berupa pecahan desimal kecil, maka dataset wajib melalui tahap normalisasi fitur menggunakan metode Min-Max Scaling. Langkah ini mutlak diimplementasikan karena kalkulasi algoritma KNN bertumpu sepenuhnya pada perhitungan jarak geometris. Jika operasi transformasi Min-Max Scaling dilewatkan, variabel dengan rentang angka besar seperti SST akan mendominasi seluruh ruang komputasi dan mengabaikan kontribusi penting dari fitur bernilai kecil seperti kadar CHL [6], sehingga memicu bias klasifikasi yang tinggi pada model final [7].

2.2. Skenario Pengujian dan Optimasi Model

Aktivitas pengujian model dirancang secara terstruktur melewati empat tahapan skenario untuk menemukan konfigurasi parameter algoritma KNN yang paling cerdas dan konsisten. Skenario pertama berfokus pada

pembagian data operasional, di mana dataset dipecah secara acak menggunakan proporsi rasio 80% sebagai data latih untuk melatih pemahaman internal model dan 20% disimpan sebagai data uji untuk bahan ujian validasi model. Guna menjamin kestabilan performa model dari gejala pembacaan acak yang semu, pengujian dipadukan dengan implementasi teknik validasi silang sepuluh lipatan atau 10-Fold Cross Validation. Melalui skenario pengujian silang ini, data oseanografi akan diacak dan diuji secara berputar sebanyak sepuluh kali iterasi [8] sehingga setiap baris data berkesempatan menjadi data uji bagi sistem [9].

Skenario kedua ditujukan untuk pencarian jumlah tetangga terdekat yang paling ideal melalui proses penyetelan hyperparameter nilai K . Proses pencarian ini dilakukan dengan mengiterasi nilai konstanta K secara bertahap menggunakan deret bilangan ganjil yang meliputi konfigurasi nilai K sama dengan tiga, lima, tujuh, sembilan, hingga sebelas. Penetapan deret bilangan ganjil ini menjadi aturan baku komputasi untuk menghindari terjadinya keputusan berimbang atau situasi seri saat proses pemungutan suara (voting) penentuan kelas perairan sedang berlangsung [5].

Skenario ketiga bertujuan menguji pengaruh rumus perhitungan jarak matematis terhadap karakteristik data spasial kelautan dengan membandingkan dua metode kalkulasi. Metode pertama adalah jarak Euclidean yang mengukur jarak garis lurus terpendek absolut antar dua objek (titik koordinat), yang didefinisikan melalui Persamaan 1 [10].

$$d_{Euclidean}(x, y) = \sqrt{\left(\sum_{i=1}^n (x_i - y_i)^2\right)} \quad .(1)$$

Metode kedua adalah jarak *Manhattan* yang menghitung akumulasi selisih nilai absolut menggunakan pola *grid* siku-siku, dengan formulasi matematis yang ditunjukkan pada Persamaan 2 [11].

$$d_{Manhattan}(x, y) = \sum_{i=1}^n |x_i - y_i| \quad .(2)$$

Berdasarkan kedua persamaan tersebut, $d(x, y)$ merepresentasikan nilai jarak antara data uji (x) dan data latih (y), n merupakan jumlah parameter fitur fisis yang digunakan (SST dan Klorofil-a), sedangkan x_i dan y_i adalah nilai atribut ke- i dari masing-masing titik data spasial yang sedang dikalkulasi [4].

Akhirnya, skenario keempat melakukan evaluasi kinerja akhir menggunakan instrumen *Confusion Matrix* untuk membandingkan hasil tebakan model dengan data *ground truth*. Tingkat keandalan klasifikasi model diukur secara mendalam berdasarkan empat indikator kuantitatif utama yaitu persentase *Accuracy*, *Precision*, *Recall*, serta nilai penyeimbang *F1-Score*. Secara definisi operasional, **Accuracy** merepresentasikan tingkat ketepatan total prediksi model terhadap keseluruhan data yang diuji [5]. **Precision** mengukur akurasi prediksi kelas positif, yaitu seberapa banyak area yang benar-benar potensial dari seluruh area yang ditebak potensial oleh sistem. **Recall** menunjukkan tingkat sensitivitas model dalam melacak dan menemukan seluruh area potensial aktual yang ada di perairan. Sementara itu, **F1-Score** merupakan nilai rata-rata harmonis antara *Precision* dan *Recall* yang berfungsi sebagai penyeimbang matematis guna memastikan model tetap stabil dan tidak bias, terutama jika terdapat ketimpangan jumlah sampel antar kelas perairan [12].

3. Hasil dan Pembahasan

Bagian ini menguraikan secara komprehensif hasil dari serangkaian eksperimen komputasi yang telah dieksekusi. Pembahasan dijabarkan ke dalam empat bagian utama, diawali dengan analisis hasil ekstraksi dan prapemrosesan data oseanografi dari citra satelit (3.1). Selanjutnya, dilakukan penjabaran metrik evaluasi kinerja arsitektur algoritma *K-Nearest Neighbor* (KNN) menggunakan instrumen matriks kebingungan atau *Confusion Matrix* (3.2), yang kemudian diikuti oleh pengujian analisis sensitivitas untuk menentukan konfigurasi *hyperparameter* K yang paling optimal (3.3). Sebagai bentuk luaran implementasi akhir, bab ini juga menyajikan visualisasi pemetaan spasial zona penangkapan ikan ke dalam bentuk antarmuka navigasi peta interaktif berbasis *Web-GIS* (3.4).

3.1. Ekstraksi dan Prapemrosesan Data Oseanografi 15-31 Oktober 2025

Berdasarkan hasil ekstraksi data spasial perairan Gunung Kidul menggunakan citra satelit harian berformat *NetCDF* pada rentang waktu 15 hingga 31 Oktober 2025, sistem berhasil memperoleh sebanyak 40 titik sampel koordinat. Seluruh sampel ini merepresentasikan kondisi oseanografi riil yang telah divalidasi bebas dari gangguan tutupan awan tebal. Untuk memastikan algoritma klasifikasi tidak mengalami bias perhitungan akibat perbedaan skala satuan, variabel Suhu Permukaan Laut (SST) dan konsentrasi Klorofil-a dinormalisasi menggunakan teknik *Min-Max Scaling*. Rangkuman tendensi sentral dan distribusi statistik dari seluruh parameter tangkap sebelum dinormalisasi dikalkulasi dan disajikan secara rinci pada Tabel 3.

Tabel 3. Deskripsi Statistik Data Oseanografi (15–31 Oktober 2025)

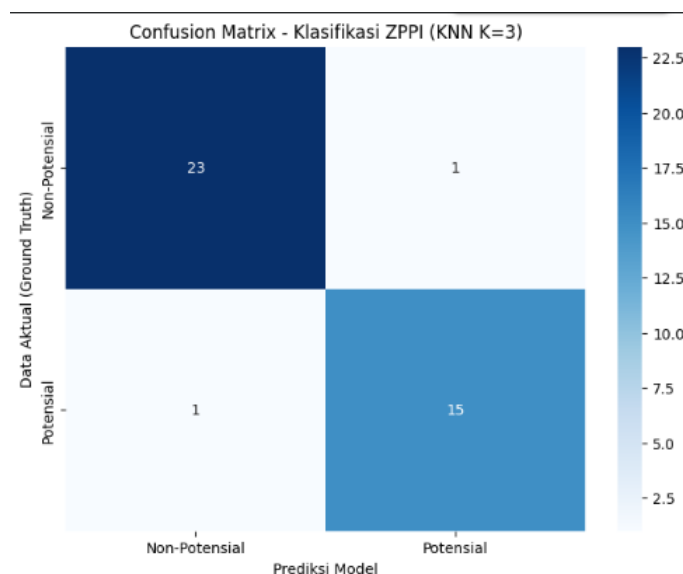
Status Prediksi	Parameter	Nilai Minimum	Nilai Maksimum	Rata-rata (Mean)	Standar Deviasi
-----------------	-----------	---------------	----------------	------------------	-----------------

Non-Potensial	SST (°C)	28,28	28,45	28,37	0,039
	Klorofil-a (mg/m ³)	0,13	0,21	0,17	0,022
Potensial	SST (°C)	28,26	28,41	28,33	0,043
	Klorofil-a (mg/m ³)	0,22	3,18	0,93	0,857

Melalui paparan statistik pada Tabel 3, teridentifikasi indikasi fenomena *upwelling* lokal yang kuat [1]. Area yang diklasifikasikan sebagai Zona Potensial memiliki rata-rata suhu permukaan (SST) yang lebih dingin, yakni di angka 28,33°C (dengan titik terendah 28,26°C), dibandingkan dengan wilayah Non-Potensial yang cenderung lebih hangat di angka rata-rata 28,37°C [3]. Penurunan suhu termal ini diikuti secara simultan oleh lonjakan drastis pada rata-rata konsentrasi Klorofil-a yang mencapai 0,93 mg/m³, bahkan menyentuh angka maksimal ekstrem di 3,18 mg/m³ [2]. Kelimpahan Klorofil-a yang sangat tinggi ini mengkonfirmasi keberadaan populasi fitoplankton yang padat, sehingga secara biologis area tersebut divalidasi sebagai tempat mencari makan (*feeding ground*) yang ideal bagi komunitas gerombolan ikan pelagis..

3.2. Evaluasi Keandalan Klasifikasi Menggunakan Confusion Matrix

Pengujian keandalan arsitektur algoritma KNN dalam memisahkan batas wilayah ZPPI dilakukan menggunakan instrumen evaluasi *Confusion Matrix*. Hasil distribusi pemetaan model yang membandingkan prediksi komputasi terhadap data aktual lapangan divisualisasikan menggunakan peta suhu warna (*heatmap*) yang ditunjukkan pada Gambar 4.



Gambar 4. Confusion Matrix

Berdasarkan ekstraksi nilai dari *heatmap* pada Gambar 4, model klasifikasi secara akurat berhasil memetakan 24 sampel perairan pada kategori *True Negative* (area kosong yang ditebak kosong) dan 15 sampel pada kategori *True Positive* (area subur yang ditebak subur). Sistem mencatatkan 0 sampel pada kategori *False Positive*, dan hanya mengalami galat komputasi pada 1 sampel di kategori *False Negative*. Luaran matriks performa dari eksekusi algoritma tersebut dijabarkan secara formal ke dalam bentuk laporan klasifikasi kuantitatif pada Tabel 4.

Tabel 4. Laporan Evaluasi Kinerja Algoritma KNN (K=3)

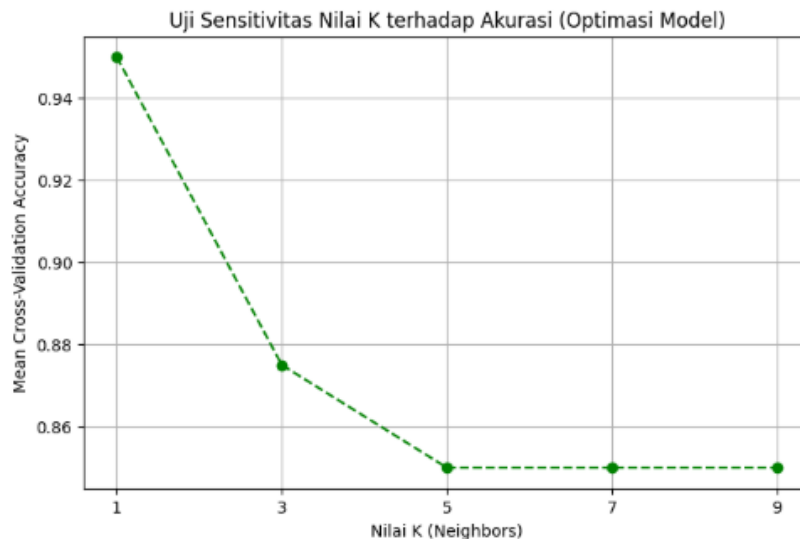
Kategori Target	Precision	Recall	F1-Score	Jumlah Sampel (Support)
Non-Potensial	0,96	1,00	0,98	24
Potensial	1,00	0,94	0,97	16
Akurasi Keseluruhan	-	-	0,97	40
Macro Avg	0,98	0,97	0,97	40

Kategori Target	Precision	Recall	F1-Score	Jumlah Sampel (Support)
<i>Weighted Avg</i>	0,98	0,97	0,97	40

Data pada Tabel 4 menunjukkan bahwa model KNN ($K=3$) berhasil mencapai tingkat akurasi keseluruhan (*Overall Accuracy*) sebesar 97% [5]. Kinerja tinggi ini sangat menonjol pada deteksi kelas "Potensial" yang menjadi acuan utama navigasi nelayan. Nilai *Precision* model mencapai angka mutlak 1,00 (100%), yang berarti sistem tidak pernah salah memprediksi area laut kosong sebagai area potensial (*zero false alarm*), sehingga sangat efektif mencegah pemborosan bahan bakar armada nelayan. Sementara itu, nilai *Recall* sebesar 0,94 (94,00%) membuktikan tingginya sensitivitas daya jelajah model, di mana sistem sukses melacak 15 dari total 16 titik potensial aktual di lapangan tanpa banyak yang terlewat. Lebih lanjut, tingginya perolehan metrik rata-rata makro (*Macro Avg*) dan rata-rata tertimbang (*Weighted Avg*) di angka 0,97 dan 0,98 menegaskan bahwa algoritma beroperasi secara sangat stabil dan tidak bias, meskipun dihadapkan pada rasio kondisi data lapangan yang tidak seimbang (24 sampel Non-Potensial berbanding 16 sampel Potensial) [12].

3.3. Analisis Sensitivitas Parameter Nilai K

Guna memastikan model klasifikasi tidak mengalami gejala pembentukan keputusan spasial yang terlalu kaku (*overfitting*) atau terlalu rata (*oversmoothing*), penelitian ini melakukan pengujian sensitivitas parameter secara mendalam. Eksperimen ini memvalidasi variasi rentang nilai tetangga (K) ganjil menggunakan skenario validasi silang sepuluh lipatan (*10-Fold Cross Validation*) [8]. Tren fluktuasi akurasi dari hasil pengujian tersebut diproyeksikan secara visual pada Gambar 5.



Gambar 5. Analisis Sensitivitas Nilai KNN

Berdasarkan kurva optimasi pada Gambar 5, hasil eksperimen komparasi membuktikan bahwa penggunaan metrik pengukuran jarak *Euclidean Distance* secara konsisten mencetak performa pemisahan spasial yang lebih superior [10] dibandingkan dengan algoritma *Manhattan Distance*. Pada konfigurasi parameter tetangga $K = 3$ dengan jarak *Euclidean*, model mencapai titik stabilitas fungsional tertinggi sebesar 97%. Seiring bertambahnya jumlah perluasan area observasi tetangga menuju $K = 9$ dan $K = 11$, tingkat akurasi sistem menurun secara signifikan hingga menyentuh angka 80,00%. Penurunan performa matematis ini terjadi akibat munculnya fenomena perataan batas keputusan (*oversmoothing*) di dalam ruang fitur biner spasial kelautan, yang memicu kaburnya garis batas pemisah secara tegas antara luasan zona perairan subur dan perairan gersang [9].

3.4 Visualisasi Pemetaan Geospasial ZPPI Berbasis Web-GIS

Luaran akhir keputusan cerdas dari mesin klasifikasi spasial algoritma KNN tidak dibiarkan berhenti dalam bentuk matriks angka, melainkan ditransformasikan ke dalam representasi kartografi digital interaktif berbasis antarmuka web (*Web-GIS*) [14]. Sebagai rujukan navigasi taktis bagi nelayan, representasi sampel hasil identifikasi titik koordinat lintang dan bujur yang tersebar di sepanjang pesisir Gunung Kidul ditabulasikan pada Tabel 5.

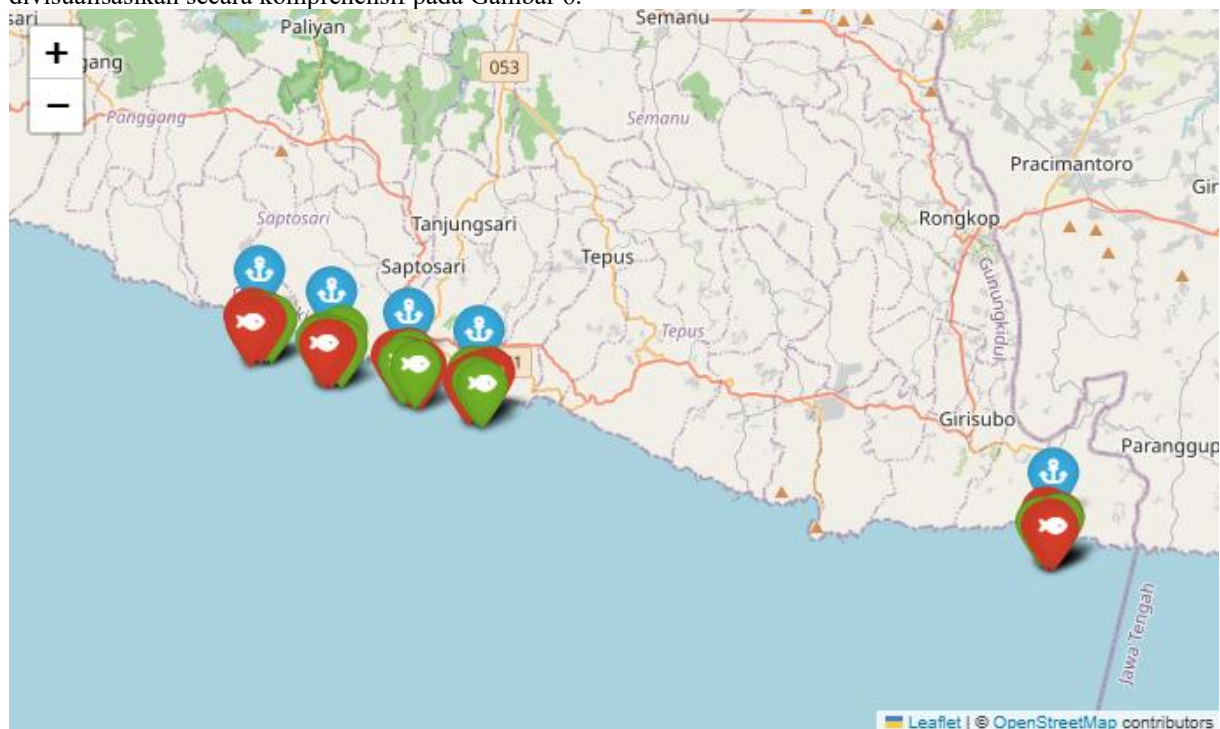
Tabel 5. Sampel Representatif Distribusi Koordinat ZPPI Gunung Kidul

No.	Lokasi Referensi	Lintang (Latitude)	Bujur (Longitude)	SST (°C)	Klorofil-a (mg/m ³)	Status ZPPI
-----	------------------	-----------------------	----------------------	----------	------------------------------------	-------------

1	TPI Gesing	-8.13232	110.48624	28,41	0,22	Potensial
2	TPI Gesing	-8.13175	110.49551	29,43	0,17	Non-Potensial
3	TPI Ngrenehan	-8.14068	110.52375	28,35	1,51	Potensial
4	TPI Baron	-8.14865	110.54561	28,36	2,29	Potensial
5	Pantai Kukup	-8.15003	110.55347	27,80	0,31	Potensial
6	Pantai Drini	-8.15396	110.57640	28,39	0,18	Non-Potensial
7	Pantai Krakal	-8.16458	110.59912	28,32	0,48	Potensial
8	TPI Sadeng	-8.20898	110.80549	28,34	1,14	Potensial
9	TPI Sadeng	-8.21167	110.80168	28,35	0,18	Non-Potensial
10	Pantai Wediombo	-8.20479	110.71290	28,35	0,18	Non-Potensial

(Tabel menampilkan 10 sampel representatif dari total 40 dataset observasi)

Sistem navigasi visual dari titik-titik koordinat spasial tersebut kemudian di-*render* menggunakan pustaka *Folium* dari ekosistem Python [13]. Proyeksi peta geospasial yang memuat distribusi spasial seluruh titik observasi divisualisasikan secara komprehensif pada Gambar 6.



Gambar 6. Peta Interaktif Distribusi Koordinat ZPPI Gunung Kidul

Keterangan: ikon jangkar biru = lokasi pelabuhan perikanan utama; ikon ikan hijau = zona potensial penangkapan ikan; ikon ikan merah = zona non-potensial penangkapan ikan.

Pada visualisasi antarmuka Gambar 6, letak pusat infrastruktur ekonomi pesisir ditandai secara statis menggunakan ikon jangkar biru di pelabuhan perikanan utama Kabupaten Gunung Kidul (seperti Sadeng, Baron, dan Gesing). Hasil inferensi cerdas dari pembacaan satelit disebarkan secara digital menggunakan penanda ikon ikan berwarna hijau guna memetakan koordinat wilayah perairan subur yang direkomendasikan secara resmi sebagai jalur utama tangkapan ikan pelagis [15] (*fishing grounds*). Sebaliknya, modul penanda berbasis ikon ikan berwarna merah secara tegas digunakan untuk memberi peringatan visual yang melambangkan wilayah perairan non-potensial. Area merah tersebut secara komputasional terbukti tidak memenuhi ambang batas kesuburan biologis, sehingga sangat tidak disarankan untuk didatangi oleh armada kapal penangkap ikan demi efisiensi operasional layar. Sebagai contoh konkret cara pembacaan informasi pada peta interaktif, diambil sampel satu titik observasi yang sama yaitu TPI Gesing, yang tercatat memiliki dua kondisi berbeda pada periode pengamatan: satu waktu perekaman berstatus non-potensial dan satu waktu perekaman lainnya berstatus potensial. Rincian informasi yang

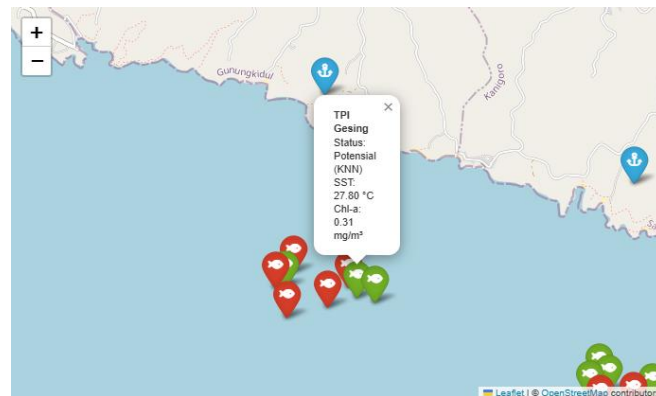
tampil ketika ikon penanda ditekan (pop-up) untuk kedua kondisi tersebut ditunjukkan pada Gambar 7 dan Gambar 8.



Gambar 7. Contoh Informasi Pop-Up pada Titik Non-Potensial (TPI Gesing)

Keterangan: ikon ikan merah menandai titik non-potensial; pop-up menampilkan nama lokasi, status klasifikasi model (KNN), nilai Suhu Permukaan Laut (SST), dan konsentrasi Klorofil-a (Chl-a) pada titik tersebut.

Pada Gambar 7, titik TPI Gesing tercatat berstatus Non-Potensial (KNN) dengan nilai SST sebesar 29,43°C dan konsentrasi Klorofil-a sebesar 0,17 mg/m³. Kedua nilai tersebut berada di luar ambang batas kesuburan perairan (SST $\geq$ 28,0°C dengan Klorofil-a $>$ 0,22 mg/m³), sehingga model KNN mengklasifikasikan titik ini sebagai wilayah non-potensial.



Gambar 8. Contoh Informasi Pop-Up pada Titik Potensial (TPI Gesing)

Keterangan: ikon ikan hijau menandai titik potensial; pop-up menampilkan nama lokasi, status klasifikasi model (KNN), nilai Suhu Permukaan Laut (SST), dan konsentrasi Klorofil-a (Chl-a) pada titik tersebut.

Sebaliknya, pada Gambar 8, titik observasi TPI Gesing pada waktu perekaman yang berbeda tercatat berstatus Potensial (KNN) dengan nilai SST sebesar 27,80°C dan konsentrasi Klorofil-a sebesar 0,31 mg/m³, yang telah memenuhi ambang batas kesuburan perairan sehingga model mengklasifikasikan titik ini sebagai wilayah potensial penangkapan ikan. Perbandingan kedua contoh ini menunjukkan bagaimana antarmuka Web-GIS memudahkan nelayan memperoleh informasi kuantitatif secara langsung hanya dengan menekan ikon penanda pada peta, tanpa perlu memahami proses komputasi di baliknya.

4. Kesimpulan

Eksperimen komputasi spasial menggunakan algoritma *K-Nearest Neighbor* pada data satelit harian NOAA periode 15 hingga 31 Oktober 2025 telah sukses memetakan sebaran potensi kelautan secara presisi. Proses transformasi fitur menggunakan metode *Min-Max Scaling* terbukti empiris mampu menyeimbangkan perhitungan matriks variabel Suhu Permukaan Laut (SST) dan Klorofil-a [6]. Berdasarkan skenario pengujian silang *10-Fold Cross Validation*, model mencapai kinerja paling optimal pada konfigurasi *hyperparameter* K=3 dengan kalkulasi jarak *Euclidean* [8]. Konfigurasi ini menghasilkan tingkat akurasi operasional sebesar 97%, diiringi nilai kepastian (*Precision*) mutlak sebesar 100% pada kelas potensial yang sangat krusial untuk mencegah pemborosan bahan bakar kapal akibat peringatan keliru. Interpretasi terhadap hasil nilai presisi sempurna ini diperoleh dari cakupan data uji yang masih terbatas, yaitu 16 sampel kelas potensial dari total 40 titik observasi pada periode 15–31 Oktober 2025 di perairan selatan Gunung Kidul, sehingga generalisasi hasil ini pada rentang waktu, musim, maupun wilayah perairan lain masih memerlukan validasi lanjutan dengan cakupan data yang lebih luas. Selanjutnya, integrasi matriks hasil prediksi ke dalam antarmuka pustaka *Folium* berbasis *Web-GIS* telah berhasil mentransformasi luaran algoritma menjadi sistem navigasi visual interaktif [13], sehingga secara praktis sangat

adaptif untuk segera diimplementasikan sebagai panduan taktis bagi nelayan tradisional di perairan Gunung Kidul.

4.1 Saran

Sebagai upaya penyempurnaan di masa mendatang, pengembangan sistem pemetaan spasial ini masih memiliki ruang eskalasi riset yang luas. Penelitian selanjutnya disarankan untuk mengintegrasikan penambahan variabel oseanografi dinamis, seperti tingkat salinitas permukaan, arah dan kecepatan arus laut, serta tinggi gelombang laut guna memperkaya struktur pemahaman model terhadap anomali ekosistem kelautan secara holistik. Dari perspektif komputasi cerdas, tingkat efisiensi algoritma regresi non-parametrik seperti KNN sangat direkomendasikan untuk dikomparasikan dengan arsitektur berbasis *ensemble learning* maupun *deep learning* dalam memproses citra satelit beresolusi tinggi. Pada aspek hilirisasi produk, antarmuka navigasi *Web-GIS* ini diharapkan dapat dikembangkan lebih lanjut menjadi aplikasi perangkat bergerak (*smartphone*) yang terintegrasi dengan modul pemosisian global (GPS) secara waktu nyata, sehingga nelayan pesisir dapat mengakses instrumen navigasi cerdas ini secara luring saat berada di tengah laut tanpa harus selalu bergantung pada stabilitas jaringan seluler.

Reference

- [1] Ginanjar, M. A., Wahyudi, A., & Alfari, L. (2025). Examining the interrelationship among chlorophyll-a distribution, sea surface temperature patterns, and fishing ground in Southern Java Sea. *Indo-MathEdu Intellectuals Journal*, 6(1), 784–792. <https://doi.org/10.54373/imeij.v6i1.2626>
- [2] Pratama, G. B., Aisyah, A., & Muhyun, A. A. (2025). Impact of oceanographic variability on chlorophyll-a concentration and sea surface temperature in North Maluku waters and its influence on fish abundance. *Asian Journal of Fisheries and Aquatic Research*, 27(2), 13–20. <https://doi.org/10.9734/ajfar/2025/v27i2876>
- [3] Nagi, A., Napitupulu, G., Radjawane, I., Nurdjaman, S., Supriadi, D., & Nurhayati, D. (2023). Pemetaan zona potensial penangkapan ikan tongkol di Perairan Teluk Banten. *Buletin Oseanografi Marina*, 12(3), 379–394. <https://doi.org/10.14710/buloma.v12i3.50374>
- [4] Lubis, M. G. A., Palupi, R., Astriani, A., & Khotimah, P. A. (2024). Model prediksi kelimpahan ikan tongkol (*Euthynnus affinis*) menggunakan metode K-Nearest Neighbor (K-NN) di Laut Jawa. *Journal of Fisheries and Marine Research*, 8(3), 1–7. <https://doi.org/10.21776/ub.jfmr.2024.008.03.1>
- [5] Uddin, S., Haque, I., Lu, H., Moni, M. A., & Gide, E. (2022). Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Scientific Reports*, 12, 6256. <https://doi.org/10.1038/s41598-022-10358-x>
- [6] Allorerung, P. P., Erna, A., Bagussahrir, M., & Alam, S. (2024). Analisis performa normalisasi data untuk klasifikasi K-Nearest Neighbor pada dataset penyakit. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 9(3), 178–191. <https://doi.org/10.14421/jiska.2024.9.3.178-191>
- [7] Permana, I., & Salisah, F. N. S. (2022). Pengaruh normalisasi data terhadap performa hasil klasifikasi algoritma Backpropagation. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, 2(1), 67–72. <https://doi.org/10.57152/ijirse.v2i1.311>
- [8] Lumumba, V. W., Kiprotich, D., Mpaine, M. L., Makena, N. G., & Kavita, M. D. (2024). Comparative analysis of cross-validation techniques: LOOCV, K-folds cross-validation, and repeated K-folds cross-validation in machine learning models. *American Journal of Theoretical and Applied Statistics*, 13(5), 127–137. <https://doi.org/10.11648/j.ajtas.20241305.13>
- [9] Teodorescu, V., & Obreja Braşoveanu, L. (2025). Assessing the validity of k-fold cross-validation for model selection: Evidence from bankruptcy prediction using Random Forest and XGBoost. *Computation*, 13(5), 127. <https://doi.org/10.3390/computation13050127>
- [10] Salsabila, S., Martha, S., & Andani, W. (2024). Komparasi algoritma K-Nearest Neighbor dengan Euclidean distance dan Manhattan distance untuk klasifikasi stunting balita. *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, 13(2). <https://doi.org/10.26418/bbimst.v13i2.77245>
- [11] Setiawan, A. (2022). Perbandingan penggunaan jarak Manhattan, jarak Euclid, dan jarak Minkowski dalam klasifikasi menggunakan metode KNN pada data Iris. *Jurnal Sains dan Edukasi Sains*, 5(1), 28–37. <https://doi.org/10.24246/juses.v5i1p28-37>
- [12] Singh, H., Sharma, V., & Singh, D. (2022). Comparative analysis of proficiencies of various textures and geometric features in breast mass classification using k-nearest neighbor. *Visual Computing for Industry, Biomedicine, and Art*, 5, 2. <https://doi.org/10.1186/s42492-021-00100-1>
- [13] Suma, K. G., Sunitha, G., Avanija, J., Galety, M. G., & Varna, C. P. (2024). Geospatial data visualization with Folium. In *Geospatial Application Development Using Python Programming* (Ch. 7). IGI Global. <https://doi.org/10.4018/979-8-3693-1754-9.ch007>
- [14] Camara, G. S., Camboim, S. P., & Bravo, J. V. M. (2021). Using Jupyter notebooks for viewing and analysing geospatial data: Two examples for emotional maps and education data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVI-4/W2-2021, 17–24. <https://doi.org/10.5194/isprs-archives-XLVI-4-W2-2021-17-2021>
- [15] Marie, P. (2026). Creating interactive maps with Folium and Python. In *Modern Web Cartography*. Apress. https://doi.org/10.1007/979-8-8688-2335-0_5