

BAB III

METODE PENELITIAN

3.1 Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksploratori-prediktif (Sugiyono, 2016). Pendekatan ini dipilih karena sesuai dengan tujuan penelitian, yaitu menganalisis tren historis produksi perikanan tangkap di PPI Binuangeun periode 2015–2024 (aspek eksploratori), dan membangun model prediksi produksi menggunakan algoritma *XGBoost* dengan mempertimbangkan variabel operasional, musiman, dan temporal (aspek prediktif). Algoritma *XGBoost* dipilih karena kemampuannya menangkap hubungan non-linier antarvariabel, yang sangat penting untuk data perikanan yang kompleks dan dipengaruhi banyak faktor. Selain itu, model ini dapat dioptimalkan dengan regularisasi L1 dan L2 serta evaluasi akurasi model secara statistik melalui metrik *MAE*, *RMSE*, *R²*, dan *MAPE*, sehingga hasil prediksi dapat dimanfaatkan secara praktis dalam perencanaan pengelolaan perikanan.

3.2 Lokasi Penelitian

3.2.1 Lokasi dan Waktu Penelitian

Penelitian ini dilaksanakan di Pangkalan Pendaratan Ikan (PPI) Binuangeun, Kabupaten Lebak, Provinsi Banten, yang secara geografis terletak pada koordinat 6°50'15" LS (Lintang Selatan) dan 105°53'11" BT (Bujur Timur). Lokasi strategis ini merupakan tempat berlabuh kapal perikanan untuk melakukan berbagai kegiatan seperti menurunkan hasil tangkapan, mengisi logistik, dan persiapan melaut kembali serta pusat aktivitas perikanan tangkap di pesisir selatan Pulau Jawa, dengan akses langsung ke perairan Samudera Hindia yang kaya sumber daya ikan (Khadijah dkk., 2019). Posisi geospasial PPI Binuangeun divisualisasikan dalam peta lokasi penelitian (Gambar 3.1),

memperlihatkan konteks wilayah studi secara komprehensif beserta batas-batas koordinatnya ($105^{\circ}41'0''$ BT hingga $106^{\circ}21'40''$ BT). Pemilihan lokasi didasarkan pada signifikansinya sebagai sentra pendaratan hasil tangkapan utama di Banten sekaligus ketersediaan data historis yang lengkap dari pengelola pelabuhan dan Dinas Kelautan-Perikanan setempat. Penelitian dilaksanakan dari 1 Juni 2025 hingga 3 Juli 2025.



Gambar 3. 1 Peta Lokasi Penelitian

3.2.2 Subjek Penelitian

Subjek penelitian ini adalah data produksi perikanan tangkap yang didaratkan di Pangkalan Pendaratan Ikan (PPI) Binuangeun, Kabupaten Lebak, Provinsi Banten. Data mencakup:

Tabel 3. 1 Variabel Penelitian

Variabel	Deskripsi	Tipe Data
Volume Produksi	Total berat hasil tangkapan per bulan (kg)	Numerik
Jumlah Kapal	Kapal yang beroperasi aktif per bulan	Numerik
Jumlah Nelayan	Nelayan aktif per bulan	Numerik
Produksi_Lag1	Volume produksi bulan sebelumnya	Numerik
Produksi_Lag3	Rata-rata volume produksi 3 bulan sebelumnya	Numerik
Produksi_Per_Kapal	Rasio produksi per kapal	Numerik

SIK UPI Kampus Serang

Variabel	Deskripsi	Tipe Data
Musim_Kemarau	Musim kemarau (Mei-Okt)	Biner (0/1)
Musim_Hujan	Musim hujan (Nov-Apr)	Biner (0/1)
Bulan_sin & Bulan_cos	Transformasi sinusoidal untuk bulan	Numerik
Tahun_Scaled	Nilai tahun yang diskalakan	Numerik
Jenis Alat Tangkap	Alat tangkap dominan (Pancing, Rampus, <i>Gillnet</i> , <i>Purse Seine</i> , Payang)	Biner (0/1)
Bulan	Periode Bulan (1-12)	Numerik
Tahun	Tahun pengamatan (2015-2024)	Numerik

Sumber: Hasil Peneliti, 2025

Data diperoleh dari berbagai sumber, termasuk Dinas Kelautan dan Perikanan Provinsi Banten, Pengelola PPI Binuangeun, dan arsip statistik bulanan pelabuhan yang dikumpulkan dari tahun 2015 hingga 2024.

3.3 Pengolahan Data

Tahapan pengolahan data dilakukan untuk mempersiapkan dataset agar sesuai dengan kebutuhan algoritma *machine learning*.

1. Penanganan Missing Values dan *Outlier*

Sebelum data digunakan untuk membangun model, langkah pertama yang dilakukan adalah mengatasi data yang hilang dan anomali (*outlier*). Data kosong diisi dengan metode *forward-fill (ffill)* untuk mengisi nilai yang hilang berdasarkan data sebelumnya, sebuah pendekatan yang efektif untuk data deret waktu yang memiliki pola berurutan (Pratama, Permanasari, Ardiyanto, & Indrayani, 2016). Setelah itu, sisa data yang masih kosong diisi dengan nilai nol (0) untuk memastikan tidak ada nilai kosong yang tersisa.

Selain itu, *outlier* pada variabel numerik, khususnya pada Volume Produksi, diidentifikasi dan dihapus menggunakan metode Interquartile Range (IQR). Pendekatan ini merupakan praktik umum dalam pra-pemrosesan data

SIK UPI Kampus Serang

untuk menjaga model dari nilai-nilai ekstrem yang dapat memengaruhi akurasi prediksi, penanganan data ini krusial untuk memastikan integritas dataset dan meningkatkan stabilitas model (Frery, 2023).

2. Rekayasa Fitur (*Feature Engineering*)

Untuk meningkatkan kinerja model, dilakukan rekayasa fitur dengan menciptakan variabel baru dari data yang sudah ada. Tahap ini krusial karena model *XGBoost* dapat belajar lebih baik dari fitur-fitur yang merepresentasikan pola data secara eksplisit (Provost dan Fawcett, 2013). Fitur-fitur yang direkayasa meliputi:

- a. Fitur *Lag*: Variabel Produksi_lag1 (produksi bulan sebelumnya) dan Produksi_lag3 (rata-rata produksi 3 bulan sebelumnya) dibuat untuk menangkap momentum historis produksi. Fitur ini sangat penting dalam analisis deret waktu karena pola produksi di bulan sebelumnya sering kali merupakan prediktor terbaik untuk produksi di bulan berikutnya (Rizal, Indah, dan Meutia, 2021).
- b. Transformasi Sinusoidal: Variabel Bulan diubah menjadi bulan_sin dan bulan_cos. Transformasi ini merepresentasikan pola musiman (seperti siklus tahunan) dalam format melingkar, yang lebih efektif bagi model *machine learning* dibandingkan dengan angka linear (1-12) (Riza, 2022).
- c. Rasio: Variabel Produksi_per_Kapal dibuat untuk merepresentasikan efisiensi penangkapan, yaitu rasio antara total produksi dengan jumlah kapal yang beroperasi.

3. Transformasi Variabel Kategorikal

Agar dapat diproses oleh algoritma *XGBoost*, variabel kategorikal perlu diubah menjadi format angka.

- a. Variabel Jenis Alat Tangkap diubah menjadi kolom-kolom biner terpisah (misalnya, Pancing, Jaring Rampus, *Gillnet*) menggunakan teknik One-Hot Encoding (Herdian dkk., 2024).
- b. Variabel Musim juga diubah menjadi kode biner, diwakili oleh variabel Musim_Hujan dan Musim_Kemarau (nilai 0 atau 1).

4. Pembagian Dataset dan Validasi Silang

Data dibagi menjadi dua bagian secara kronologis agar pola waktu atau temporal tetap terjaga. Sebanyak 80% data, yang mencakup periode 2015–2021, digunakan sebagai data latih (training set), sementara sisanya 20%, yang mencakup periode 2022–2024, digunakan sebagai data uji (testing set).

Dalam tahap pencarian parameter terbaik (*hyperparameter tuning*) untuk model *XGBoost*, digunakan metode *RandomizedSearchCV* dengan validasi silang (*k-fold cross-validation*). Teknik ini bermanfaat untuk mengurangi risiko bias dan *overfitting*, serta memastikan model bekerja secara konsisten, sejalan dengan temuan (Arifin, Ariawan, Rosalia, Lukman, & Tufailah, 2022).

3.4 Batasan Penelitian

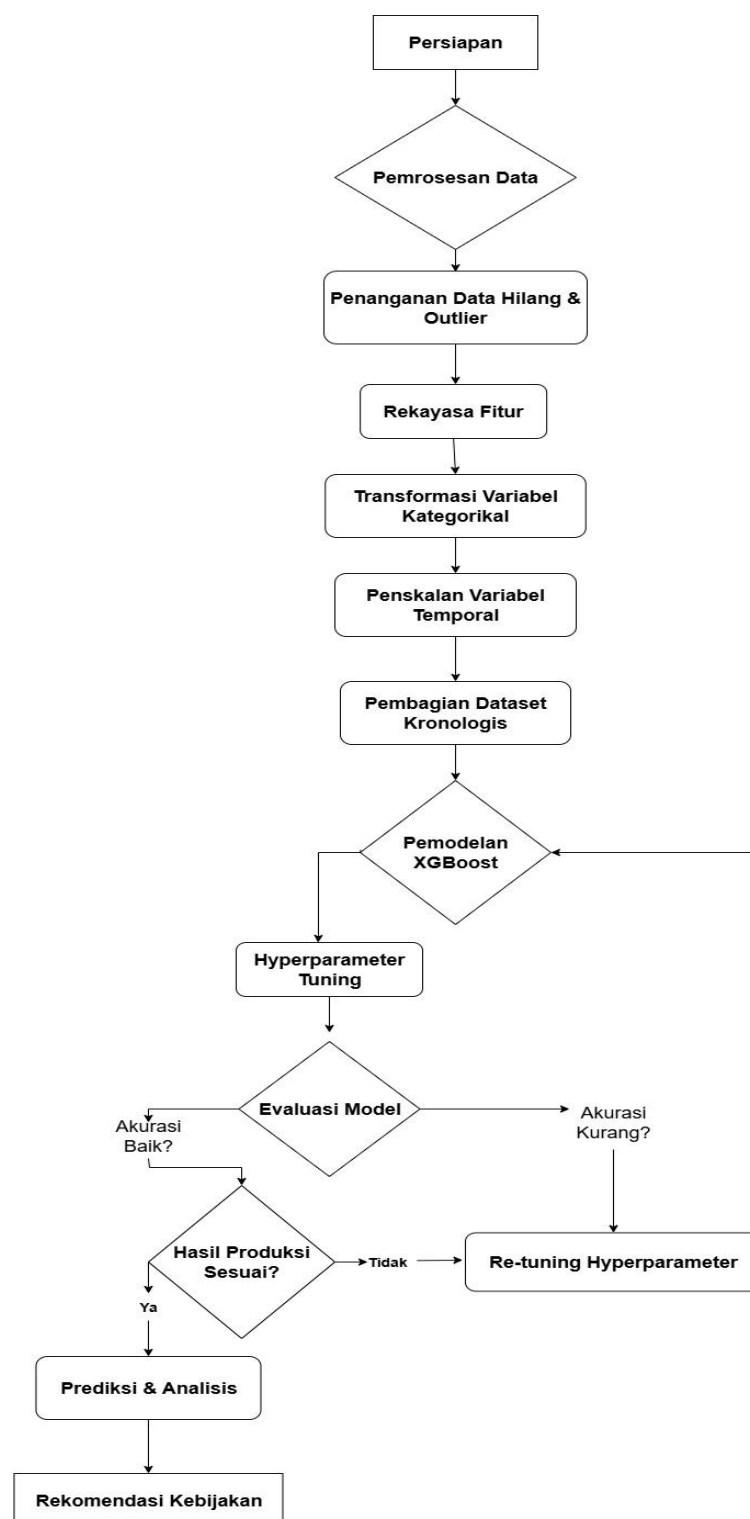
Penelitian ini memiliki beberapa batasan yang perlu diperhatikan agar hasilnya dapat dipahami secara tepat, yaitu:

1. Penelitian hanya menggunakan data historis periode 2015–2024 yang diperoleh dari PPI Binuangeun, sehingga hasil prediksi terbatas pada ruang lingkup data tersebut.
2. Faktor iklim global ENSO (El Niño–Southern Oscillation) tidak dimasukkan ke dalam model, sehingga variabel iklim eksternal tidak memengaruhi hasil prediksi.
3. Model yang digunakan terbatas pada *XGBoost* dengan regularisasi L1 dan L2, tanpa melakukan perbandingan dengan algoritma lain seperti *ARIMA*, *LSTM*, atau *Random Forest*.

SIK UPI Kampus Serang

4. Hasil prediksi hanya berlaku pada kondisi data historis yang digunakan, sehingga apabila terjadi perubahan signifikan di masa depan, misalnya perubahan kebijakan, iklim, atau infrastruktur perikanan, maka hasil prediksi dapat berbeda.

3.5 Prosedur Penelitian



Gambar 3. 2 Diagram Alur Metode Penelitian

SIK UPI Kampus Serang

Gambar 3.2 menunjukkan alur penelitian, dimulai dari pengumpulan data produksi perikanan, dilanjutkan dengan pembersihan dan persiapan data, pemodelan menggunakan algoritma XGBoost, evaluasi model dengan metrik MAE, RMSE, R^2 , dan MAPE, hingga menghasilkan prediksi produksi tahun 2025–2027. Prosedur penelitian mengikuti tahapan:

1. Persiapan

Melakukan studi literatur terkait algoritma *XGBoost* dan dinamika perikanan di PPI Binuangeun. Mengumpulkan data sekunder produksi perikanan tangkap bulanan periode 2015–2024 dari Dinas Kelautan dan Perikanan Provinsi Banten.

2. Pemrosesan Data

Tahap ini meliputi beberapa langkah penting untuk mempersiapkan data:

- a. Penanganan Data Hilang dan *Outlier*: Mengisi data kosong pada variabel deret waktu menggunakan metode forward fill (ffill) diikuti dengan pengisian nilai nol (0) (Pratama, 2016). Selain itu, nilai-nilai ekstrem (*outlier*) pada variabel Produksi diidentifikasi dan dihapus menggunakan metode Interquartile Range (IQR) (Frery, 2023).
- b. Rekayasa Fitur (*Feature Engineering*): Menciptakan variabel baru untuk meningkatkan kemampuan model dalam menangkap pola data yang kompleks. Fitur-fitur yang dibuat meliputi:
 - 1) Fitur lag: Produksi_lag1 (produksi bulan sebelumnya) dan Produksi_lag3 (rata-rata produksi 3 bulan sebelumnya) (Rizal, 2021).
 - 2) Transformasi siklus: bulan_sin dan bulan_cos untuk merepresentasikan pola musiman tahunan (Riza, 2022).
 - 3) Fitur rasio: Produksi_per_Kapal untuk mengukur efisiensi penangkapan.
- c. Transformasi Variabel Kategorikal: Mengubah variabel Jenis Alat Tangkap menjadi format biner melalui one-hot encoding (Herdian dkk.,

SIK UPI Kampus Serang

2024) dan variabel Musim menjadi kode biner (Musim_Hujan, Musim_Kemarau).

- d. Penskalaan Variabel Temporal: Mengubah variabel Tahun menjadi Tahun_scaled dengan mengurangi nilai tahun minimum untuk membantu model menangkap tren jangka panjang secara lebih efektif (Arifin dkk., 2022).
- e. Pembagian Dataset: Data dibagi secara kronologis menjadi 80% data latih (2015–2021) dan 20% data uji (2022–2024) untuk menjaga integritas pola waktu.

3. Pemodelan *XGBoost*:

Tahap pemodelan menggunakan algoritma *Extreme Gradient Boosting (XGBoost)* (Chen dan Guestrin, 2016). Model dibangun menggunakan kelas *XGBRegressor* dengan seluruh variabel hasil pemrosesan data sebagai input. Untuk mengoptimalkan kinerja model, dilakukan *hyperparameter tuning* menggunakan metode *RandomizedSearchCV* dengan validasi silang (*k-fold cross-validation*). Proses ini juga menyertakan parameter regularisasi L1 (*reg_alpha*) dan L2 (*reg_lambda*) untuk mencegah *overfitting* dan meningkatkan generalisasi model (Probst dkk., 2019).

4. Evaluasi Model:

Uji akurasi model dilakukan dengan membandingkan hasil prediksi (*y_pred*) pada data uji dengan data aktual (*y_test*) menggunakan metrik-metrik berikut:

a. *MAE (Mean Absolute Error)*

MAE menghitung rata-rata selisih absolut antara nilai prediksi dan aktual. Metrik ini memberikan gambaran kesalahan prediksi dalam satuan asli data dan bersifat mudah diinterpretasikan (Suryanto & Muqtadir, 2019).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \dots\dots\dots (1)$$

MAE menunjukkan rata-rata selisih antara hasil prediksi dan data

SIK UPI Kampus Serang

sebenarnya dalam satuan kilogram.

b. *RMSE (Root Mean Squared Error)*

RMSE adalah akar kuadrat dari rata-rata kuadrat selisih antara nilai aktual dan prediksi. Nilainya berada pada satuan yang sama dengan variabel target, sehingga memudahkan interpretasi. *RMSE* sensitif terhadap *outlier* sehingga cocok digunakan untuk mengukur kesalahan yang besar pada data tertentu (Neri, 2024).

$$RMSE = \sqrt{\left(\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2\right)} \dots\dots\dots (2)$$

RMSE mengukur seberapa besar kesalahan prediksi, dengan memberikan bobot lebih besar pada kesalahan yang besar.

c. *R² (Koefisien Determinasi)*

R² menunjukkan proporsi variasi data yang dapat dijelaskan oleh model. Nilainya berkisar antara 0 hingga 1, di mana nilai yang lebih tinggi menunjukkan kemampuan model yang lebih baik dalam menjelaskan data (Neri, 2024).

$$R^2 = 1 - \frac{SSR}{SST} \dots\dots\dots (3)$$

dengan *SSR* adalah jumlah kuadrat residu, dan *SST* adalah jumlah kuadrat total. *R²* menggambarkan seberapa baik model menjelaskan variasi data, dengan nilai mendekati 1 berarti model semakin baik.

d. *MAPE (Mean Absolute Percentage Error)*

MAPE mengukur rata-rata persentase kesalahan prediksi terhadap nilai aktual (Lewis, 1982). Nilai *MAPE* yang lebih kecil menunjukkan tingkat akurasi yang lebih tinggi. Rumus:

$$MAPE = \frac{100\%}{N} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \dots\dots\dots (4)$$

MAPE menampilkan kesalahan dalam bentuk persen, sehingga lebih

SIK UPI Kampus Serang

mudah dipahami pembuat kebijakan. *Mean Absolute Percentage Error (MAPE)* adalah salah satu ukuran yang sering digunakan untuk menilai tingkat akurasi hasil prediksi. Nilai *MAPE* yang lebih kecil menunjukkan hasil prediksi yang lebih mendekati nilai sebenarnya. Tabel berikut menyajikan interpretasi umum nilai *MAPE* sebagaimana dikemukakan oleh (Lewis, 1982).

Tabel 3. 2 Kriteria Akurasi Berdasarkan Nilai *MAPE*

MAPE	Interpretasi
<10%	Sangat Akurat
10-20%	Baik/Akurat
20-50%	Cukup/Wajar
>50%	Kurang Akurat

Sumber: (Lewis, 1982)

Tabel di atas menjelaskan ukuran ketepatan model prediksi berdasarkan nilai *MAPE*, kalau nilai *MAPE* kurang dari 10%, berarti prediksinya sangat akurat karena hampir sama dengan data sebenarnya, jika nilainya berada di antara 10 sampai 20 persen, hasil prediksinya masih tergolong baik atau akurat dan bisa dipercaya, apabila *MAPE* ada di kisaran 20 sampai 50 persen, maka prediksinya dianggap cukup wajar, meskipun ada selisih yang lumayan terasa, sedangkan sekiranya lebih dari 50%, prediksi tersebut sudah kurang akurat dan sulit dijadikan pegangan (Lewis, 1982).

5. Prediksi dan Analisis

Tahap terakhir adalah melakukan prediksi volume produksi perikanan tangkap untuk periode 2025–2027. Proyeksi ini dilakukan dengan memproyeksikan fitur input masa depan menggunakan tren historis seperti *CAGR (Compound Annual Growth Rate)* dan indeks musiman. Hasil prediksi ini kemudian dianalisis dan dibandingkan dengan tren historis untuk melihat potensi peningkatan atau penurunan produksi di masa mendatang, yang menjadi dasar perumusan rekomendasi kebijakan.

SIK UPI Kampus Serang