Evaluasi dan Perbandingan Kinerja Model Regresi *Machine Learning* dalam Prediksi Produksi Panen Ikan Lele Tahunan di Indonesia

1st Isnaeni Fatmawati
Program Studi S1 Teknik Informatika
Universitas Telkom Purwokerto
Purwokerto, Indonesia
isnaenifatmawati@student.telkomuniversity.ac.id

2nd Paradise
Program Studi S1 Teknik Informatika
Universitas Telkom Purwokerto
Purwokerto, Indonesia
paradise@telkomuniversity.ac.id

3rd Mahazam Afrad
Program Studi S1 Teknik Informatika
Universitas Telkom Purwokerto
Purwokerto, Indonesia
mahazama@telkomuniversity.ac.id

***Abstrak* — Produksi ikan lele merupakan salah satu komoditas penting dalam sektor perikanan budidaya air tawar di Indonesia. Namun, produksi ikan lele tahunan masih mengalami fluktuasi serta perbedaan yang cukup signifikan antar provinsi, sehingga perencanaan produksi dan distribusi sering kali belum optimal. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja beberapa model regresi berbasis *machine learning* dalam memprediksi produksi panen ikan lele tahunan pada tingkat provinsi di Indonesia. Data yang digunakan merupakan data sekunder produksi ikan lele periode 2017–2024 yang bersumber dari Badan Pusat Statistik dan Kementerian Kelautan dan Perikanan Republik Indonesia. Metode penelitian meliputi *preprocessing* data, pembentukan fitur keterkaitan waktu berupa fitur *lag* dan fitur perubahan produksi, serta pembagian data menggunakan pendekatan *time-aware split* untuk menjaga urutan temporal data. Model yang digunakan dalam penelitian ini meliputi *Ridge Regression*, *Random Forest Regressor*, *Histogram-based Gradient Boosting Regressor*, dan *Light Gradient Boosting Machine (LightGBM)*. Evaluasi kinerja model dilakukan menggunakan metrik *RMSE*, *MAE*, *MAPE*, dan koefisien determinasi (R^2). Hasil penelitian menunjukkan bahwa *Ridge Regression* memberikan performa paling stabil dan optimal dalam memprediksi produksi ikan lele tahunan pada tingkat provinsi.**

***Kata kunci* — *machine learning*, *regresi*, *time series*, produksi ikan lele, prediksi produksi, perikanan budidaya**

I. PENDAHULUAN

Produksi ikan lele merupakan salah satu komoditas strategis dalam sektor perikanan budidaya air tawar di Indonesia karena memiliki nilai ekonomi yang tinggi, mudah dibudidayakan, serta menjadi sumber protein hewani yang terjangkau bagi masyarakat. Tingginya permintaan pasar terhadap ikan lele mendorong peningkatan aktivitas budidaya di berbagai daerah, baik pada skala rumah tangga maupun komersial. Namun demikian, produksi ikan lele tahunan di Indonesia masih menunjukkan fluktuasi dari tahun ke tahun serta perbedaan yang cukup signifikan antar provinsi, yang dapat berdampak pada ketidakseimbangan pasokan dan distribusi di pasar nasional [1], [2].

Fluktuasi produksi ikan lele dipengaruhi oleh berbagai faktor, seperti kondisi lingkungan, harga pakan, serangan penyakit, serta perbedaan karakteristik budidaya antar wilayah. Dalam praktiknya, perencanaan produksi ikan lele masih banyak mengandalkan pengalaman dan intuisi pembudidaya, sehingga sering kali kurang didukung oleh analisis data historis yang sistematis [3]. Kondisi ini menyebabkan prediksi produksi panen belum optimal dan berpotensi menimbulkan ketidakstabilan pasokan serta harga di pasar.

Pendekatan prediksi berbasis data historis menjadi salah satu solusi untuk mendukung perencanaan produksi yang lebih akurat. Metode statistik konvensional memiliki keterbatasan dalam menangkap pola data yang kompleks dan bersifat nonlinier, terutama pada data deret waktu dengan karakteristik fluktuatif [4]. Oleh karena itu, pendekatan *machine learning* semakin banyak digunakan dalam berbagai penelitian prediksi produksi di sektor agrikultur dan perikanan karena kemampuannya dalam memodelkan hubungan nonlinier serta pola temporal secara lebih adaptif [5], [6].

Beberapa penelitian sebelumnya menunjukkan bahwa model regresi berbasis *machine learning*, seperti *Random Forest*, *Gradient Boosting*, dan metode regresi teratur (*regularized regression*), mampu memberikan performa yang lebih baik dibandingkan metode konvensional dalam memprediksi hasil produksi berbasis data historis [7], [8]. Meskipun demikian, pemilihan model yang paling sesuai sangat bergantung pada karakteristik data yang digunakan, sehingga diperlukan analisis perbandingan kinerja antar model secara sistematis.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja beberapa model regresi *machine learning* dalam memprediksi produksi panen ikan lele tahunan pada tingkat provinsi di Indonesia dengan memanfaatkan data historis produksi. Model yang digunakan dalam penelitian ini meliputi *Ridge Regression*, *Random Forest Regressor*,

Histogram-based Gradient Boosting Regressor, dan *Light Gradient Boosting Machine* (*LightGBM*). Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam pemilihan model prediksi yang optimal serta menjadi dasar pendukung pengambilan keputusan dalam perencanaan produksi perikanan budidaya berbasis data di Indonesia.

II. KAJIAN TEORI

A. Ikan Lele

Ikan lele merupakan salah satu komoditas unggulan perikanan budidaya air tawar di Indonesia karena memiliki tingkat adaptasi yang tinggi, waktu panen yang relatif singkat, serta permintaan pasar yang stabil. Produksi ikan lele tersebar di berbagai provinsi dengan karakteristik budidaya yang berbeda-beda, sehingga menghasilkan variasi produksi tahunan yang cukup signifikan antar wilayah. Data produksi ikan lele umumnya disajikan dalam bentuk deret waktu tahunan yang dipengaruhi oleh faktor lingkungan, manajemen budidaya, dan kondisi ekonomi [9], [10].

B. *Machine Learning*

Machine learning merupakan pendekatan komputasi yang memungkinkan sistem untuk mempelajari pola dari data historis dan menghasilkan prediksi tanpa pemodelan eksplisit berbasis aturan tertentu. Dalam konteks prediksi produksi, *machine learning* banyak digunakan karena kemampuannya dalam menangkap hubungan nonlinier dan kompleks antar variabel. Berbagai penelitian menunjukkan bahwa pendekatan *machine learning* mampu memberikan hasil prediksi yang lebih akurat dibandingkan metode statistik konvensional, khususnya pada data produksi yang bersifat fluktuatif [11], [12].

C. Regresi dalam *Machine Learning*

Regresi merupakan salah satu pendekatan *supervised learning* yang bertujuan untuk memprediksi nilai kontinu berdasarkan pola data historis. Model regresi dalam *machine learning* banyak digunakan pada permasalahan prediksi produksi karena dapat memodelkan hubungan antara variabel *input* dan *output* secara fleksibel. Beberapa model regresi yang umum digunakan antara lain regresi linear teratur (*regularized regression*), regresi berbasis pohon keputusan, serta metode *ensemble* berbasis boosting [7].

D. Ridge Regression

Ridge Regression merupakan pengembangan dari regresi linear yang menerapkan regularisasi L2 untuk mengurangi masalah multikolinearitas dan *overfitting* pada data. Penambahan parameter penalti pada fungsi objektif memungkinkan model untuk menghasilkan estimasi parameter yang lebih stabil, terutama pada data dengan jumlah fitur yang saling berkorelasi. *Ridge Regression* sering digunakan pada data *time series* dengan ukuran data relatif terbatas karena memiliki kompleksitas model yang rendah dan interpretabilitas yang baik [13].

E. Random Forest Regressor

Random Forest Regressor merupakan metode *ensemble learning* yang membangun banyak pohon keputusan secara acak dan menggabungkan hasil prediksinya. Model ini mampu menangkap hubungan nonlinier dan interaksi kompleks antar variabel tanpa memerlukan asumsi distribusi data tertentu. Dalam berbagai penelitian, *Random Forest* terbukti efektif dalam permasalahan prediksi produksi karena memiliki ketahanan terhadap *noise* dan *overfitting* [14].

F. *Gradient Boosting* dan *LightGBM*

Histogram-based Gradient Boosting Regressor dan *Light Gradient Boosting Machine* (*LightGBM*) merupakan metode *boosting* yang membangun model secara bertahap dengan meminimalkan kesalahan prediksi dari model sebelumnya. *LightGBM* menggunakan pendekatan berbasis histogram sehingga lebih efisien dalam komputasi dan memori. Metode *boosting* banyak digunakan dalam prediksi produksi karena mampu memberikan performa tinggi pada data dengan pola kompleks dan nonlinier [15], [16].

G. Data *Time Series* dan *Feature Engineering*

Data produksi ikan lele tahunan termasuk ke dalam kategori data *time series*, yaitu data yang memiliki ketergantungan terhadap urutan waktu. Untuk meningkatkan performa prediksi, diperlukan proses *feature engineering* yang mempertimbangkan keterkaitan temporal, seperti pembentukan fitur *lag* dan fitur perubahan (*difference*). Pendekatan ini memungkinkan model *machine learning* untuk mempelajari pola tren dan fluktuasi produksi dari waktu ke waktu secara lebih efektif [17].

H. Evaluasi Kinerja Model Regresi

Evaluasi kinerja model regresi bertujuan untuk mengukur tingkat akurasi prediksi yang dihasilkan. Beberapa metrik evaluasi yang umum digunakan meliputi *Root Mean Square Error* (*RMSE*), *Mean Absolute Error* (*MAE*), *Mean Absolute Percentage Error* (*MAPE*), dan koefisien determinasi (R^2). Penggunaan beberapa metrik secara bersamaan memberikan gambaran yang lebih komprehensif terhadap performa model dalam memprediksi data produksi [6], [7].

III. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membangun dan mengevaluasi model prediksi produksi panen ikan lele tahunan berbasis *machine learning*. Tahapan penelitian disusun secara sistematis mulai dari pengumpulan data, *preprocessing*, *feature engineering*, pembagian data berbasis waktu, pelatihan dan optimasi model, hingga evaluasi kinerja model prediksi.

A. Data dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder produksi ikan lele tahunan pada tingkat provinsi di Indonesia dengan rentang waktu 2017–2024. Data diperoleh dari instansi resmi pemerintah, yaitu Badan Pusat Statistik (BPS) untuk periode 2017–2018 dan Kementerian Kelautan dan Perikanan Republik Indonesia (KKP) untuk periode 2019–2024, sehingga sumber data yang digunakan memiliki tingkat kredibilitas dan konsistensi yang tinggi.

Data produksi ikan lele disajikan dalam satuan ton per tahun untuk masing-masing provinsi, yang mencerminkan jumlah hasil panen ikan lele pada periode tertentu. Data tersebut dimodelkan sebagai data deret waktu panel (panel *time series*) yang memiliki dimensi waktu dan dimensi

wilayah. Karakteristik data ini memungkinkan dilakukan analisis pola produksi ikan lele baik secara temporal antar tahun maupun secara spasial antar provinsi. Selain itu, struktur data panel memberikan fleksibilitas dalam pembentukan fitur keterkaitan waktu, seperti fitur *lag* dan perubahan produksi, yang penting untuk meningkatkan kemampuan model *machine learning* dalam menangkap dinamika produksi ikan lele tahunan di Indonesia.

TABEL 1
(Variabel Dataset Penelitian)

No	Nama Variabel	Satuan	Keterangan
1	Tahun	Tahun	Periode waktu data produksi
2	Provinsi	Kategori	Wilayah administrasi
3	Produksi Lele	Ton/Kg	Jumlah produksi aktual (target)

Tabel 1 menyajikan daftar variabel yang digunakan dalam penelitian untuk memprediksi produksi ikan lele tahunan. Variabel yang digunakan meliputi informasi waktu, wilayah, variabel target produksi ikan lele, serta fitur keterkaitan waktu berupa fitur *lag* dan perubahan produksi. Seluruh variabel dinyatakan dalam satuan ton per kilogram (Ton/Kg) dan disusun untuk merepresentasikan karakteristik data deret waktu produksi ikan lele pada tingkat provinsi.

B. *Preprocessing* Data

Tahap *preprocessing* data dilakukan untuk memastikan kualitas dan konsistensi data sebelum digunakan dalam pemodelan. Proses ini meliputi standarisasi format data, penyesuaian tipe data numerik, penanganan nilai hilang, serta penghapusan data yang tidak lengkap. Selanjutnya, data diurutkan berdasarkan provinsi dan tahun untuk menjaga urutan temporal setiap wilayah. Tahapan ini penting karena penelitian menggunakan pendekatan *time series,* sehingga ketidakteraturan urutan waktu dapat memengaruhi hasil prediksi.

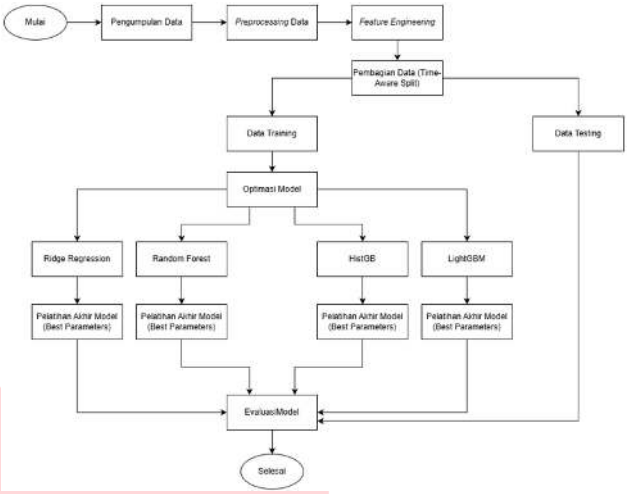
Untuk memperjelas tahapan *preprocessing* yang diterapkan pada data produksi ikan lele, rangkuman proses *preprocessing* yang dilakukan pada penelitian ini disajikan dalam Tabel 2.

TABEL 2
(Tahapan *Preprocessing* Data)

No	Tahapan	Deskripsi
1	Tandarisasi Data	Memastikan struktur dan penamaan atribut konsisten agar siap diproses lebih lanjut.
2	Eksplorasi Data (EDA)	Menganalisis karakteristik awal data, meliputi jumlah data, rentang tahun, provinsi, dan nilai hilang.
3	Data *Cleaning*	Menangani nilai tidak valid, menyesuaikan tipe data numerik, dan merapikan data kategorikal.
4	Seleksi dan Pengurutan Data	Menghapus data yang tidak lengkap dan mengurutkan data berdasarkan provinsi dan waktu.

Tabel 2 merangkum tahapan *preprocessing* data yang dilakukan sebelum proses pemodelan, meliputi penyesuaian format dan tipe data, penanganan nilai hilang, serta pengurutan data berdasarkan provinsi dan tahun. Tahapan tersebut bertujuan untuk memastikan data tersusun secara konsisten dan sesuai dengan karakteristik data deret waktu sehingga dapat digunakan secara optimal pada proses pelatihan model.

Alur penelitian pada penelitian ini dirancang secara sistematis untuk memastikan setiap tahapan pemodelan dilakukan secara terstruktur dan berurutan.



GAMBAR 1
(Diagram Alir Penelitian)

Gambar 3.2 menunjukkan alur penelitian yang digunakan dalam penelitian ini. Tahapan penelitian dimulai dari pengumpulan data produksi ikan lele, dilanjutkan dengan *preprocessing* data, pembentukan fitur keterkaitan waktu, pembagian data berbasis urutan waktu, pelatihan dan optimasi model regresi *machine learning*, serta evaluasi kinerja model prediksi.

C. *Feature Engineering*

Feature engineering dilakukan untuk menangkap keterkaitan temporal pada data produksi ikan lele tahunan. Pada penelitian ini dibentuk fitur *lag* dan fitur perubahan produksi. Fitur *lag* merepresentasikan nilai produksi pada satu hingga tiga tahun sebelumnya (*lag-1*, *lag-2*, dan *lag-3*), sedangkan fitur perubahan produksi (*diff-1*) menggambarkan selisih produksi antar periode. Pembentukan fitur dilakukan secara terpisah untuk setiap provinsi agar pola temporal masing-masing wilayah tetap terjaga. Data yang tidak memiliki informasi *lag* lengkap tidak digunakan dalam proses pemodelan.

TABEL 3
(Data Setelah Pembentukan Fitur *Lag)*

Provinsi	Tahun	Produksi Lele	Lag-1	Lag-2	Lag3
Aceh	2020	11635	12779	10997	13141
Aceh	2021	6610	11635	12779	10997
Aceh	2022	7812	6610	11635	12779
Aceh	2023	6669	7812	6610	11635
Aceh	2024	5234	6669	7812	6610

Berdasarkan Tabel 3, terlihat bahwa nilai produksi ikan lele pada tahun berjalan dilengkapi dengan informasi produksi pada tahun-tahun sebelumnya. Keberadaan fitur *lag* ini memungkinkan model untuk mempelajari pola historis produksi serta keterkaitan antar periode waktu dalam satu provinsi. Dengan memanfaatkan informasi tersebut, model diharapkan mampu menangkap tren dan fluktuasi produksi

secara lebih akurat sehingga meningkatkan kualitas prediksi produksi ikan lele tahunan.

D. Pembagian Data (*Time-Aware Split*)

Pembagian data dilakukan menggunakan pendekatan *time-aware split*, yaitu pembagian data berdasarkan urutan waktu, bukan secara acak. Data pada tahun-tahun awal digunakan sebagai data latih, sedangkan data pada tahun terakhir digunakan sebagai data uji. Pendekatan ini bertujuan untuk menghindari kebocoran data (data *leakage*) dan memastikan bahwa proses evaluasi mencerminkan kondisi prediksi yang realistis.

TABEL 4
(Ringkasan Pembagian Data Berdasarkan *Time-Aware Split)*

Komponen	Keterangan
Rentang data latih	2020-2023
Tahun data uji	2024
Jumlah data latih	135 baris
Jumlah data uji	33 baris
Jumlah fitur input	7 fitur

Berdasarkan Tabel 4, sebagian besar data dimanfaatkan sebagai data latih untuk memberikan informasi historis yang memadai bagi model dalam mempelajari pola produksi ikan lele. Sementara itu, data uji yang berasal dari tahun 2024 digunakan untuk mengevaluasi kemampuan model dalam melakukan prediksi pada periode waktu yang benar-benar baru. Dengan skema pembagian ini, model dilatih menggunakan data historis dan diuji pada kondisi yang menyerupai penerapan nyata, sehingga hasil evaluasi yang diperoleh lebih realistis dan dapat mencerminkan kemampuan generalisasi model.

E. Pelatihan dan Optimasi Model

Penelitian ini membandingkan empat model regresi *machine learning*, yaitu *Ridge Regression*, *Random Forest Regressor*, *Histogram-based Gradient Boosting Regressor*, dan *Light Gradient Boosting Machine* (*LightGBM*). Proses pelatihan model dilakukan menggunakan pipeline terintegrasi dan optimasi parameter menggunakan *Randomized Search Cross-Validation* dengan skema *expanding window*. Pendekatan ini memungkinkan model dilatih secara bertahap mengikuti pertambahan data historis sehingga lebih sesuai dengan karakteristik data *time series*.

Untuk model *Ridge Regression*, proses pelatihan dilakukan dengan meminimalkan fungsi objektif yang menggabungkan kesalahan prediksi dan penalti regularisasi L2, yang secara matematis dirumuskan sebagai berikut:

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - x_i^T \beta\right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \qquad (1)$$

Keterangan:

n = jumlah data observasi

p = jumlah variabel prediktor

y_i = nilai aktual pada observasi ke- i

x_i = vektor fitur pada observasi ke- i

λ = parameter regulasi L2

$\sum_{j=1}^{p} \beta_j^2$ = penalti kuadrat koefisien regresi

Selain penentuan fungsi objektif, optimasi parameter dilakukan pada seluruh model regresi yang digunakan dalam penelitian ini. Parameter yang dioptimasi meliputi parameter utama yang berpengaruh terhadap kompleksitas model dan kinerja prediksi. Tabel 5 menyajikan ringkasan parameter optimasi yang digunakan pada masing-masing model regresi *machine learning*.

TABEL 5
(Parameter Optimasi Model Regresi)

Model	Parameter	Rentang Nilai
Ridge Regression	α (alpha)	10^{-3} – 10^{3}
Random Forest Regressor	Jumlah Pohon	300, 600, 900
	Kedalaman Maksimum	None, 6, 10, 14, 18
	Minimum Sampel Split	2, 5, 10
	Minimum Sampel Daun	1, 2, 4
	Fitur maksimum	*sqrt*, 0.5, 0.8
Histogram-based Gradient Boosting	*Learning Rate*	0.03, 0.05, 0.1, 0.2
	Kedalaman Maksimum	3, 5, 7, None
	Jumlah Iterasi	300, 600, 900
	Minimum Sampem Daun	10, 20, 30, 50
	Regulasi L2	0.0, 0.1, 1.0
Light Gradient Boosting Machine (LightGBM)	*Learning Rate*	0.03, 0.05, 0.1
	Jumlah Daun	15, 31, 63
	Kedalaman Maksimun	-1, 4, 6, 8
	Minimum sampel anak	2, 5, 10, 20
	Subsampling	0.07, 0.85, 1.0
	Column sampling	0.07, 0.85, 1.0
	Regulasi L1	0.0, 0.1
	Regulasi L2	0.0, 0.1, 1.0

Tabel 4 menyajikan parameter yang dioptimasi pada masing-masing model regresi *machine learning* beserta rentang nilai yang digunakan dalam proses *Randomized Search Cross-Validation*. Pemilihan parameter difokuskan pada pengaturan kompleksitas model dan kemampuan generalisasi, sehingga diharapkan menghasilkan performa prediksi yang optimal pada data produksi ikan lele tahunan.

F. Evaluasi dan Kinerja Model

Evaluasi kinerja model dilakukan menggunakan beberapa metrik regresi, yaitu *Root Mean Square Error* (*RMSE*), *Mean Absolute Error* (*MAE*), *Mean Absolute Percentage Error* (*MAPE*), dan koefisien determinasi (R^2). Penggunaan beberapa metrik bertujuan untuk memberikan gambaran yang komprehensif mengenai tingkat kesalahan prediksi dan kemampuan model dalam menjelaskan variasi data produksi ikan lele tahunan.

RMSE digunakan untuk mengukur rata-rata kesalahan prediksi dengan memberikan penalti lebih besar pada kesalahan berukuran besar. Secara matematis, *RMSE* dirumuskan sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i}^{n}\left(\tilde{y}^i - y^i\right)^2} \qquad (2)$$

Keterangan:

n = jumlah data uji

$\tilde{y}^i$ = nilai prediksi ke- i

y^i = nilai aktual ke- i

$(\tilde{y}^i - y^i)^2$ = kuadrat dari selisih nilai prediksi dan aktual

Selain *RMSE*, koefisien determinasi (R^2) digunakan untuk mengukur kemampuan model dalam menjelaskan variasi data aktual. Nilai R^2 berada pada rentang 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan kemampuan model yang semakin baik. Secara matematis, R^2 dirumuskan sebagai berikut:

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(X_i - Y_i)^2}{\sum_{i=1}^{m}(\bar{Y} - Y_i)^2} \quad (3)$$

Keterangan:

m = jumlah data observasi

X_i = nilai prediksi ke- i

Y_i = nilai aktual ke- i

$\bar{Y}$ = rata-rata dari seluruh nilai aktual

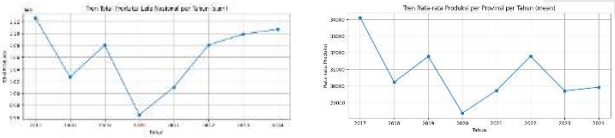
Metrik *MAE* dan *MAPE* digunakan sebagai metrik pendukung untuk mengukur kesalahan prediksi secara absolut dan persentase. *MAE* mengukur rata-rata selisih absolut antara nilai aktual dan nilai prediksi, sedangkan *MAPE* mengukur kesalahan prediksi dalam bentuk persentase, sehingga memudahkan interpretasi kinerja model secara relatif.

IV. HASIL DAN PEMBAHASAN

Implementasi model regresi *machine learning* untuk prediksi produksi ikan lele tahunan pada tingkat provinsi telah dilakukan sesuai dengan tahapan metodologi yang diusulkan. Evaluasi dilakukan secara komprehensif untuk menilai kinerja model berdasarkan beberapa metrik regresi, serta menganalisis kemampuan model dalam memodelkan pola produksi ikan lele tahunan. Hasil eksperimen dan analisis visualisasi prediksi pada data produksi ikan lele menunjukkan perbedaan performa antar model regresi yang digunakan, yang selanjutnya dibahas secara lebih rinci pada subbab berikutnya.

A. Karakteristik Data Produksi Ikan Lele

Data produksi ikan lele tahunan dianalisis terlebih dahulu untuk mengidentifikasi pola umum, tren waktu, dan perbedaan produksi antar provinsi. Analisis ini diperlukan untuk memberikan konteks terhadap hasil pemodelan yang dilakukan pada tahap selanjutnya.



Gambar 4.1 (a) Grafik tren total produksi ikan lele nasional per tahun
(b) Grafik tren rata-rata produksi ikan lele per provinsi

Grafik tren produksi nasional menunjukkan bahwa produksi ikan lele cenderung mengalami fluktuasi dari tahun ke tahun dengan kecenderungan peningkatan pada periode tertentu. Sementara itu, grafik rata-rata produksi per provinsi memperlihatkan adanya perbedaan tingkat produksi antar wilayah, yang mengindikasikan bahwa karakteristik produksi ikan lele tidak bersifat homogen dan dipengaruhi oleh kondisi masing-masing provinsi.

B. Hasil Pelatihan dan Evaluasi Model Regresi

Setelah data diproses dan model dilatih, dilakukan evaluasi kinerja untuk membandingkan kemampuan masing-masing model regresi dalam memprediksi produksi ikan lele tahunan. Evaluasi ini bertujuan untuk mengidentifikasi model dengan performa prediksi terbaik.

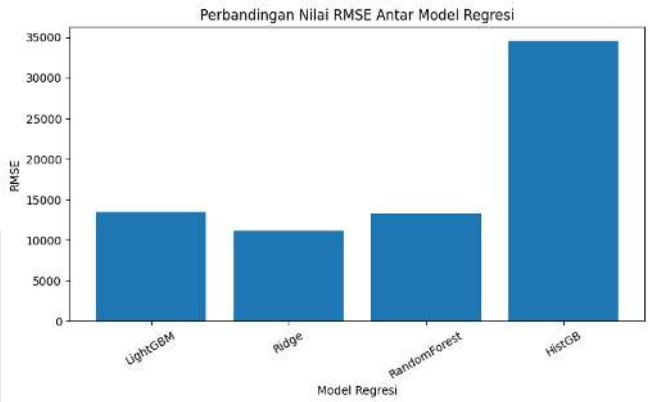
TABEL 6
(Perbandingan Kinerja Model (*RMSE*, *MAE*, *MAPE*, R^2))

Model	*Best Cv RMSE*	*RMSE*	*MAE*	*MAPE*	R^2
Ridge	**8774.540**	**7828.543**	**5247.583**	**587.371**	**0.985**
LightGBM	9107.610	9559.786	4830.864	49.981	0.978
Randon Forest	10577.478	9207.503	4468.754	57.480	0.980
HistGB	24584.587	20373.361	8342.426	61.149	0.903

Berdasarkan tabel evaluasi, *Ridge Regression* menghasilkan nilai *RMSE*, *MAE*, dan *MAPE* yang lebih rendah dibandingkan *Random Forest*, *Histogram-based Gradient Boosting*, dan *LightGBM*. Selain itu, nilai koefisien determinasi (R^2) yang dihasilkan *Ridge Regression* relatif lebih stabil, yang menunjukkan kemampuan model tersebut dalam menjelaskan variasi data produksi ikan lele tahunan dengan lebih baik.

C. Visualisasi Perbandingan Kinerja Model

Untuk memperjelas perbedaan performa antar model, hasil evaluasi divisualisasikan dalam bentuk grafik. Visualisasi ini memudahkan interpretasi perbedaan tingkat kesalahan prediksi antar model regresi.

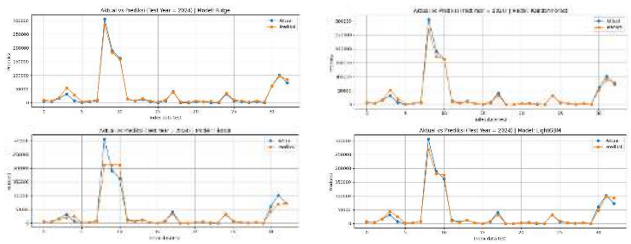


GAMBAR 2
(Grafik perbandingan nilai RMSE antar model regresi)

Grafik perbandingan *RMSE* menunjukkan bahwa *Ridge Regression* memiliki tingkat kesalahan prediksi paling rendah dibandingkan model lainnya. Model berbasis pohon keputusan dan *boosting* menunjukkan variasi kesalahan yang lebih besar, yang mengindikasikan performa yang kurang stabil pada data produksi ikan lele tahunan.

D. Hasil Prediksi Produksi Menggunakan Model Terbaik

Berdasarkan hasil evaluasi dan visualisasi kinerja, *Ridge Regression* dipilih sebagai model terbaik. Selanjutnya, dilakukan analisis terhadap hasil prediksi produksi ikan lele yang dihasilkan oleh model tersebut.



GAMBAR 3
(Hasil Prediksi Produksi Ikan Lele Tahun 2024 (a) *Ridgde Regression* (b) *Random Forest Regressor* (c) *Histogram-Based Gradient Boosting* (d) *LightGBM)*

Grafik perbandingan nilai aktual dan prediksi menunjukkan bahwa hasil prediksi *Ridge Regression* mampu mengikuti pola tren produksi ikan lele tahunan dengan cukup baik. Meskipun terdapat selisih pada beberapa periode, secara umum model dapat merepresentasikan fluktuasi produksi secara konsisten.

E. Pembahasan Hasil

Hasil penelitian menunjukkan bahwa *Ridge Regression* lebih sesuai digunakan untuk memprediksi produksi ikan lele tahunan dibandingkan model regresi nonlinier lainnya. Hal ini disebabkan oleh karakteristik data produksi tahunan yang memiliki jumlah observasi relatif terbatas serta hubungan antar variabel yang cenderung linear. Penerapan regularisasi L2 pada *Ridge Regression* berperan dalam mengurangi risiko *overfitting*, sehingga model mampu menghasilkan prediksi yang lebih stabil.

Sebaliknya, model *Random Forest*, *Histogram-based Gradient Boosting*, dan *LightGBM* memiliki tingkat kompleksitas yang lebih tinggi dan umumnya membutuhkan jumlah data yang lebih besar untuk mencapai performa optimal. Pada data produksi tahunan dengan skala terbatas, model-model tersebut cenderung menunjukkan performa yang kurang konsisten dibandingkan *Ridge Regression*.

Penelitian ini masih memiliki ruang untuk pengembangan lebih lanjut. Penelitian selanjutnya dapat mempertimbangkan penggunaan data dengan resolusi waktu yang lebih tinggi, seperti data bulanan atau triwulanan, serta penambahan variabel eksternal yang berpotensi memengaruhi produksi ikan lele, seperti faktor cuaca dan harga pakan. Selain itu, penerapan pendekatan *deep learning* dapat dieksplorasi untuk menangkap pola nonlinier yang lebih kompleks.

V. KESIMPULAN

Penelitian ini menunjukkan bahwa pendekatan regresi berbasis *machine learning* dapat digunakan secara efektif untuk memprediksi produksi ikan lele tahunan pada tingkat provinsi di Indonesia. Berdasarkan hasil evaluasi kinerja menggunakan metrik *RMSE*, *MAE*, *MAPE*, dan koefisien determinasi (R^2), *Ridge Regression* memberikan performa yang paling stabil dan akurat dibandingkan *Random Forest*, *Histogram-based Gradient Boosting*, dan *Light Gradient Boosting Machine*. Keunggulan *Ridge Regression* dipengaruhi oleh karakteristik data produksi tahunan yang memiliki jumlah observasi terbatas dan pola hubungan yang cenderung linear, sehingga regularisasi L2 mampu mengurangi risiko *overfitting*. Model prediksi yang dihasilkan diharapkan dapat menjadi alat bantu berbasis data dalam mendukung perencanaan produksi perikanan budidaya, khususnya ikan lele, di Indonesia. Hasil penelitian ini juga menunjukkan bahwa pemilihan model yang sesuai dengan karakteristik data memiliki peran penting dalam meningkatkan akurasi prediksi produksi tahunan.

REFERENSI

[1] A. A. Rahma, R. S. Nurlaela, A. Meilani, Z. P. Saryono, and A. D. Pajrin, "Ikan Sebagai Sumber Protein dan Gizi Berkualitas Tinggi Bagi Kesehatan Tubuh Manusia," *Karimah Tauhid*, vol. 3, no. 3, pp. 3132–3142, 2024, doi: 10.30997/karimahtauhid.v3i3.12341.

[2] "Goal 2: Zero Hunger - United Nations Sustainable Development." Accessed: May 21, 2025. [Online]. Available: https://www.un.org/sustainabledevelopment/hunger/

[3] "Produksi Perikanan Budidaya Menurut Komoditas Utama - Tabel Statistik - Badan Pusat Statistik Indonesia." Accessed: May 21, 2025. [Online]. Available: https://www.bps.go.id/id/statistics-table/2/MTUxMyMy/produksi-perikanan-budidaya-menurut-komoditas-utama.html?utm_source=chatgpt.com

[4] S. A. Ananda and K. Sa'diyah, "Analisis Ekonomi Prarancangan Pabrik Kimia Pembuatan Pakan Ikan Lele Dari Maggot Dengan Kapasitas 20.000 Ton/Tahun," *DISTILAT: Jurnal Teknologi Separasi*, vol. 10, no. 1, pp. 188–196, 2024, doi: 10.33795/distilat.v10i1.4206.

[5] L. Adeliya Nababan, "Jurnal JPILKOM (Jurnal Penelitian Ilmu Komputer) Prediksi Produksi Ikan Lele Menggunakan Algoritma Backpropagation dalam Jaringan Syaraf Tiruan," Online, 2025.

[6] F. W. Tina, N. Afsarimanesh, A. Nag, and M. E. E. Alahi, "Integrating AIoT Technologies in Aquaculture: A Systematic Review," *Future Internet*, vol. 17, no. 5, pp. 1–35, 2025, doi: 10.3390/fi17050199.

[7] A. A. Rahmadani *et al.*, "Data-Driven Predictions of Fish Production: Applying Regression Methods in Aquaculture," *International Journal of Frontier Technology and Engineering (IJFTE)*, vol. 02, no. 02, pp. 65–78, 2024, [Online]. Available: http://jurnal.polinema.ac.id/index.php/IJFTE/index

[8] L. F. Rahman, M. Marufuzzaman, L. Alam, M. A. Bari, U. R. Sumaila, and L. M. Sidek, "Developing an ensembled machine learning prediction model for marine fish and aquaculture production," *Sustainability (Switzerland)*, vol. 13, no. 16, Aug. 2021, doi: 10.3390/su13169124.

[9] EMILDA SAKINA and A.HAIDAR MIRZA, "PREDIKSI HASIL PRODUKSI IKAN LELE MENGGUNAKAN MACHINE LEARNING (Studi Kasus Dinas Perikanan Kabupaten Muara Enim)," *Jurnal INSTEK (Informatika Sains dan Teknologi)*, vol. 9, no. 1, pp. 55–64, 2024, doi: 10.24252/instek.v9i1.46406.

[10] M. Nahak, "Bab Ii Tinjauan Pustaka Dan Landasan Teori," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 21–

25, 2017, [Online]. Available: http://www.elsevier.com/locate/scp

[11] "Machine Learning Tutorial - GeeksforGeeks." Accessed: Dec. 18, 2025. [Online]. Available: https://www.geeksforgeeks.org/machine-learning/machine-learning/

[12] A. Darmawan Sidik, A. Ansawarman, K. Kunci, J. Kendaraan Bermotor, M. Regresi, and F. Jalan, "Prediksi Jumlah Kendaraan Bermotor Menggunakan Machine Learning," *Formosa Journal of Multidisciplinary Research (FJMR)*, vol. 1, no. 3, pp. 559–568, 2022, doi: 10.55927.

[13] "Ridge Regression - GeeksforGeeks." Accessed: Dec. 18, 2025. [Online]. Available: https://www.geeksforgeeks.org/machine-learning/what-is-ridge-regression/

[14] H. A. Salman, A. Kalakech, and A. Steiti, "Random Forest Algorithm Overview," Dec. 15, 2024, *Mesopotamian Academic Press*. doi: 10.58496/BJML/2024/007.

[15] H. Deng, Y. Zhou, L. Wang, and C. Zhang, "Ensemble learning for the early prediction of neonatal jaundice with genetic features," *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, Dec. 2021, doi: 10.1186/s12911-021-01701-9.

[16] "Light GBM vs XGBOOST: Which algorithm takes the crown." Accessed: Aug. 10, 2025. [Online]. Available: https://www.analyticsvidhya.com/blog/2017/06/which-algorithm-takes-the-crown-light-gbm-vs-xgboost/

[17] Y. Setiawan, T. Tarno, and P. Kartikasari, "Prediksi Harga Jual Kakao Dengan Metode Long Short-Term Memory Menggunakan Metode Optimasi Root Mean Square Propagation Dan Adaptive Moment Estimation Dilengkapi Gui Rshiny," *Jurnal Gaussian*, vol. 11, no. 1, pp. 99–107, 2022, doi: 10.14710/j.gauss.v11i1.33994.