

OPTIMASI MODEL PREDIKSI EXTREME GRADIENT BOOSTING DENGAN GENETIC ALGORITHM UNTUK PRODUKSI DAN PRODUKTIVITAS PADI

¹⁾ Kadek Ardy Wirakusuma, ²⁾ Ni Putu Novita Puspa Dewi, ³⁾ Kadek Yota Ernanda Aryanto
^{1,2,3)}Program Studi Ilmu Komputer, Jurusan Teknik dan Informatika, Fakultas Teknik dan Kejuruan,
Universitas Pendidikan Ganesha
¹⁾ ardy.wirakusuma@student.undiksha.ac.id,

INFO ARTIKEL	ABSTRAK
<p>Riwayat Artikel : Diterima : 20 Oktober 2025 Disetujui : 25 Oktober 2025</p> <p>Kata Kunci : Produksi Padi, Produktivitas Padi, Genetic Algorithm, XGBoost, GA- XGBoost</p>	<p>Produksi padi di Kabupaten Buleleng meningkat dari tahun 2021 hingga 2023, namun produktivitas justru menurun hingga 8%. Kondisi ini berpotensi mengganggu stabilitas pasokan beras di tengah pertumbuhan penduduk sebesar 4.59% per tahun. Diperlukan pendekatan prediktif berbasis kecerdasan buatan untuk memodelkan hubungan kompleks antar variabel pertanian. XGBoost dipilih karena kemampuannya dalam menangkap pola non-linear dan sering digunakan dalam analisis data pertanian, sementara Genetic Algorithm (GA) digunakan untuk menentukan kombinasi hiperparameter optimal guna meningkatkan performa model. Model XGBoost tanpa optimasi diterapkan sebagai pembanding untuk mengevaluasi efektivitas pendekatan <i>hybrid</i>. Hasil analisis menunjukkan bahwa optimasi hiperparameter berpengaruh signifikan terhadap hasil prediksi. Model GA-XGBoost menunjukkan kinerja yang lebih baik dibandingkan model standar, dengan penurunan nilai MAPE sebesar 2.99 poin persentase pada prediksi produksi padi (28.48% menjadi 25.49%) dan 0.21 poin persentase pada prediksi produktivitas padi (2.52% menjadi 2.31%).</p>
ARTICLE INFO	ABSTRACT
<p>Article History : Received : Oct 20, 2025 Accepted : Oct 25, 2025</p> <p>Keywords: Rice Production, Rice Productivity, Genetic Algorithm, XGBoost, GA- XGBoost</p>	<p><i>Rice production in Buleleng Regency increased from 2021 to 2023, however, productivity declined by up to 8%. This condition has the potential to disrupt rice supply stability amid a population growth rate of 4.59% per year. A predictive approach based on artificial intelligence is required to model the complex relationships among agricultural variables. XGBoost was chosen for its ability to capture non-linear patterns and its frequent use in agricultural data analysis, while the Genetic Algorithm (GA) was employed to determine the optimal combination of hyperparameters to improve model performance. The non-optimized XGBoost model was used as a baseline to evaluate the effectiveness of the hybrid approach. The analysis results show that hyperparameter optimization significantly influences prediction outcomes. The GA-XGBoost model achieved lower error rates, with a reduction in MAPE of 2.99 percentage points for rice production predictions (28.48% to 25.49%) and 0.21 percentage points for productivity predictions (2.52% to 2.31%) compared to the standard model</i></p>

1. PENDAHULUAN

Tanaman padi merupakan komoditas pertanian yang sangat vital bagi masyarakat Indonesia karena menjadi sumber utama bahan pangan sehari-hari. Ketersediaan beras perlu dijaga agar mampu memenuhi permintaan yang terus meningkat (Rahayu et al., 2023). Produksi padi yang stabil dan mencukupi berperan penting dalam mencegah terjadinya defisit serta menekan potensi kerawanan pangan, termasuk di wilayah Kabupaten Buleleng. Berdasarkan data Badan Pusat Statistik (BPS) Kabupaten Buleleng (2024), produksi padi mengalami peningkatan sebesar 8,408 ton Gabah Kering Giling (GKG) dalam periode 2021 hingga 2023, yaitu dari 82,975 ton menjadi 91,383 ton.

Peningkatan produksi padi dipengaruhi oleh bertambahnya luas panen, yang mencapai 16,885 hektar pada tahun 2023, meningkat sekitar 2,773 hektar (19%) dibandingkan dua tahun sebelumnya. Namun, hal ini tidak sejalan dengan produktivitas padi sendiri. Pada periode yang sama, produktivitas padi menurun sekitar 8%, dari 5.88 ton per hektar menjadi 5.41 ton per hektar. Penurunan ini berpotensi mengganggu kestabilan pasokan beras, terutama apabila perluasan lahan tidak mampu mengimbangi peningkatan kebutuhan akibat pertumbuhan penduduk (Zhang and Jia, 2021). Secara ideal, produksi dan produktivitas seharusnya meningkat bersamaan. Fakta bahwa produktivitas menurun mengindikasikan adanya faktor-faktor tertentu yang memengaruhi, seperti penurunan kualitas lahan, penerapan teknologi pertanian yang belum optimal, atau dampak perubahan iklim (Sriartha et al., 2023).

Penurunan produktivitas yang berkelanjutan dapat mengancam kestabilan produksi jangka panjang, terutama dengan laju pertumbuhan penduduk Kabupaten Buleleng yang mencapai 4.59% per tahun (BPS Kabupaten Buleleng, 2024). Untuk itu, diperlukan pendekatan prediktif berbasis teknologi yang mampu memantau sekaligus memperkirakan produksi dan produktivitas padi secara lebih akurat. Salah satu metode yang dapat diterapkan adalah pemanfaatan kecerdasan buatan, khususnya *machine learning*, dalam sistem prediksi pertanian. Model *Extreme Gradient Boosting* (XGBoost) menjadi salah satu algoritma yang banyak digunakan karena kemampuannya

mengenali pola non-linear serta kinerjanya yang unggul dalam analisis dan prediksi hasil pertanian (Swain et al., 2024). Penelitian oleh (Tabassum et al., 2024) membuktikan bahwa XGBoost lebih akurat dibanding *Random Forest* dalam memprediksi variasi kedalaman gerusan pada tanggul pancang. Sementara itu, studi oleh (Noorunnahar et al., 2023) menunjukkan XGBoost mengungguli ARIMA dalam memprediksi produksi beras di Bangladesh, dengan nilai MAPE sebesar 5.38% dibandingkan 7.23% pada ARIMA.

Dalam penerapannya, meskipun XGBoost telah terbukti unggul dalam berbagai penelitian, performanya sangat dipengaruhi oleh pemilihan hiperparameter yang tepat. Oleh karena itu, penerapan algoritma optimasi menjadi krusial untuk menemukan kombinasi hiperparameter yang paling efektif dan efisien (Dewi et al., 2025; Handayani Putri et al., 2023). Beragam algoritma optimasi telah digunakan untuk meningkatkan performa model prediksi, seperti *Fruit Fly Optimization Algorithm* (Dewi et al., 2021; Putu et al., 2021), *Flower Pollination Algorithm* (Pascima and Putrama, 2021), dan *Genetic Algorithm* (Dedy et al., 2023; Samantaray et al., 2024). Selain itu, (Dewi et al., 2025) juga menunjukkan bahwa penerapan optimasi berbasis *Bayesian Optimization* dengan *Gaussian Process* dapat secara signifikan meningkatkan performa model LSTM dan GRU, sekaligus mempercepat proses pencarian hiperparameter yang optimal. Di antara semua metode tersebut, *Genetic Algorithm* (GA) dikenal unggul karena kemampuannya menyeimbangkan antara eksplorasi global dan eksploitasi lokal dalam proses pencarian solusi (Tran & Nguyen, 2024).

Sejumlah penelitian telah membuktikan efektivitas *Genetic Algorithm* (GA) dalam meningkatkan kinerja model prediksi. Studi (Zhang & Jia, 2021) menunjukkan bahwa penerapan GA pada XGBoost mampu menurunkan nilai MAPE dari 2.22% menjadi 1.62% dalam kasus prediksi beban listrik. Sementara itu, (Jiang et al., 2021) menunjukkan bahwa kombinasi GA dengan XGBoost-SVR menghasilkan nilai MSE yang lebih rendah dibandingkan model standar tanpa optimasi. Selain itu, (Aydin & Gülsün, 2024) menemukan bahwa XGBoost yang dioptimasi dengan GA

memiliki performa lebih baik dibandingkan XGBoost dengan *Grey Wolf Optimizer* (GWO) dalam memprediksi jumlah kasus COVID-19.

Dalam penelitian ini, digunakan model *hybrid* XGBoost dan GA sebagai model prediksi. GA berperan dalam menentukan nilai hiperparameter optimal bagi XGBoost. Sebagai pembanding, model XGBoost tanpa optimasi turut diterapkan untuk menilai efektivitas dari pendekatan *hybrid* yang diusulkan.

2. METODE

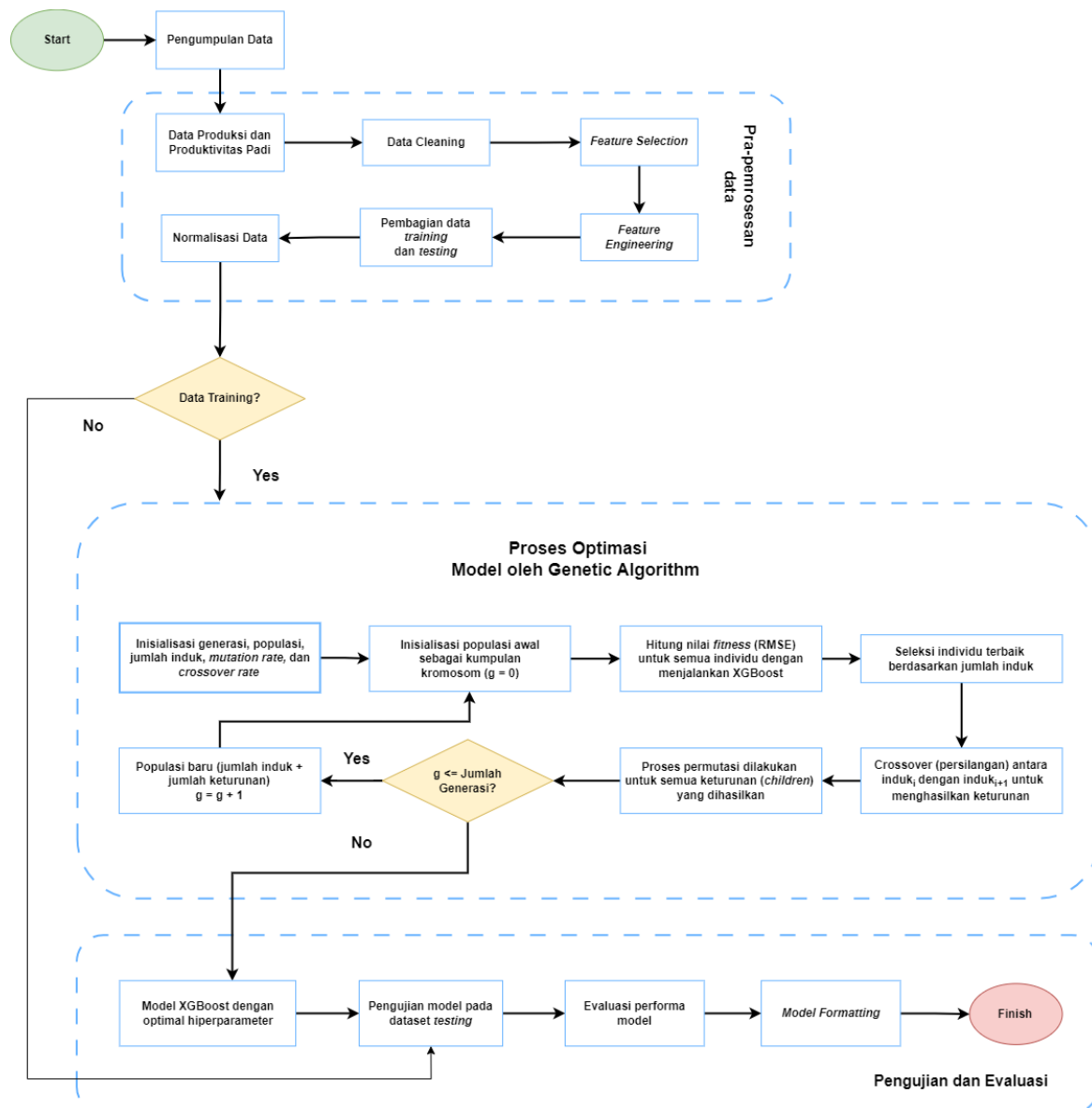
2.1. Wilayah Penelitian

Penelitian ini memanfaatkan data produksi dan produktivitas padi yang diperoleh langsung dari Badan Pusat Statistik (BPS) Kabupaten Buleleng. Kabupaten Buleleng sendiri memiliki luas wilayah sekitar 1,365 km² atau sekitar

24.25% dari total luas Provinsi Bali. Sebagian besar penduduknya menggantungkan mata pencaharian pada sektor pertanian, peternakan, perikanan, serta perkebunan, yang menjadi tulang punggung ekonomi daerah.

2.2. Alur Penelitian

Alur penelitian dibagi menjadi tiga sub-proses, yaitu pra-pemrosesan data, optimasi model XGBoost dengan *Genetic Algorithm*, serta pengujian dan evaluasi model. Setiap proses yang ditandai dengan garis putus-putus pada Gambar 1 menunjukkan masing-masing subproses yang terlibat.



Gambar 1. Alur Penelitian

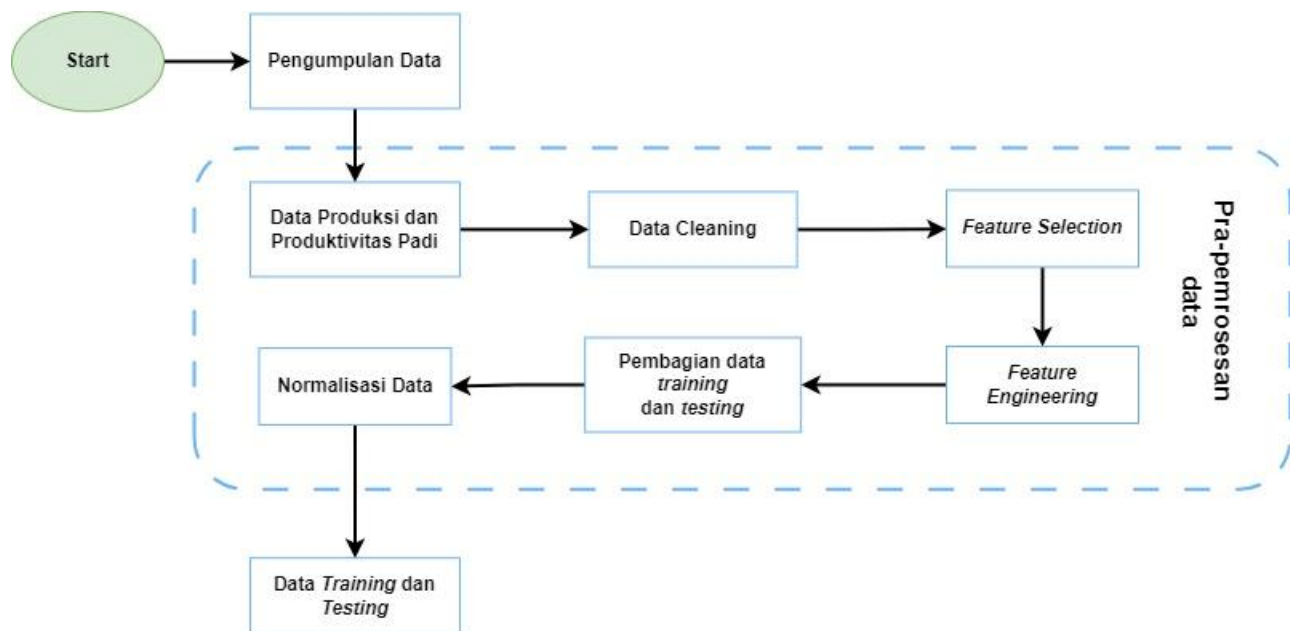
2.2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh langsung dari Badan Pusat Statistik (BPS) Kabupaten Buleleng dan mencakup dua variabel utama, yaitu luas panen (hektar) dan produksi padi (ton). Sedangkan data produktivitas padi dihitung sebagai rasio antara jumlah produksi dan luas panen. Seluruh data yang diperoleh kemudian disimpan dalam format .xlsx (*Excel Open XML Spreadsheet*)

untuk selanjutnya diolah dan dianalisis lebih lanjut. Untuk lebih jelasnya Gambar 2 menjelaskan setiap proses, khususnya pengumpulan data hingga normalisasi data.

2.2.2. Pra-pemrosesan Data

Setelah seluruh data terkumpul, dilakukan tahapan pra-pemrosesan yang meliputi *data cleaning*, *feature selection*, *feature engineering*, pembagian data, dan normalisasi data.



Gambar 2. Pra-pemrosesan Data

Data cleaning mencakup pengecekan konsistensi waktu, interpolasi data yang hilang, serta deteksi dan penanganan *outlier* untuk memastikan kualitas dataset. Selanjutnya, *feature selection* dilakukan menggunakan *feature importance* hingga *time-step* $t - 12$ untuk menangkap pola musiman dan tren tahunan, serta analisis *Cross Correlation Function* (CCF) untuk melihat hubungan antara luas panen dan produksi untuk berbagai *time-step*. Berdasarkan hasil tersebut, tahap *feature engineering* dilakukan melalui *reframing* atau *windowing*, yaitu mengubah data deret waktu menjadi bentuk *cross-sectional* dengan akumulasi nilai historis untuk memprediksi periode berikutnya. *Window size* ditentukan dari hasil analisis *feature importance* dan CCF.

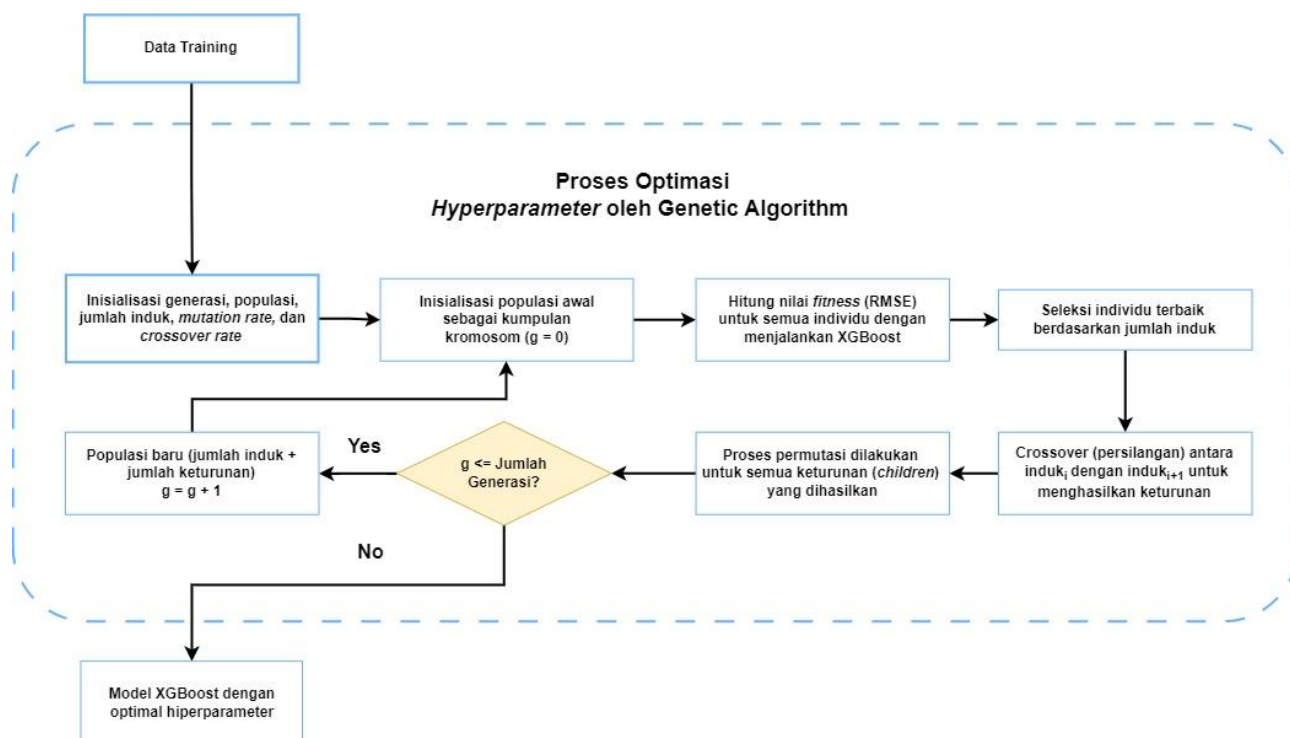
Setelah fitur terbentuk, dataset dibagi menjadi 80% data latih dan 20% data uji dalam format *sliding window*. Tahap akhir adalah normalisasi menggunakan metode *Min-Max Normalization* untuk menyamakan skala variabel produksi dan luas panen.

2.2.3. Optimasi Model

Dalam proses optimasi, setiap hiperparameter memiliki perannya masing-masing dalam mengontrol kompleksitas, generalisasi, dan efisiensi model. Pemilihan nilai hiperparameter yang tepat sangat penting untuk memperoleh performa model yang optimal. Oleh karena itu, diperlukan tahapan sistematis untuk mengeksplorasi kombinasi terbaik dari hiperparameter yang tersedia.

Gambar 3 memperlihatkan alur proses optimasi, di mana data train digunakan untuk menemukan

kombinasi hiperparameter terbaik oleh GA sebelum dilakukan evaluasi pada data *testing*.



Gambar 3. Optimasi Model

Tabel 1 memuat nama, tipe data, dan deskripsi hiperparameter yang digunakan. Selanjutnya ditentukan rentang nilai hiperparameter *Genetic Algorithm* (GA), seperti jumlah generasi, induk, populasi, *crossover rate*, dan *mutation rate*. GA kemudian digunakan untuk mengoptimalkan hiperparameter XGBoost, dengan hasil akhir berupa kombinasi terbaik beserta nilai galat yang diusulkan yaitu *Root Mean Squared Error* (RMSE).

Tabel 1. Hiperparameter Model XGBoost

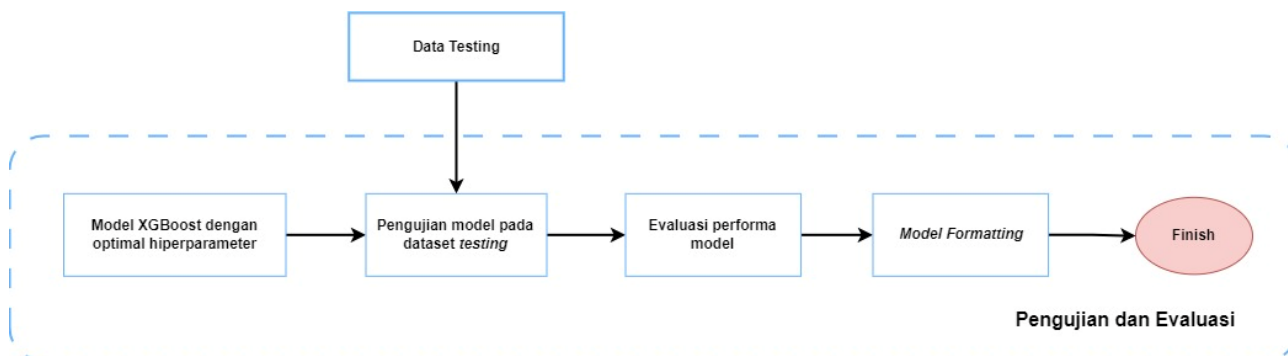
Hiperparameter	Tipe Data	Deskripsi
N_estimators	Integer	Jumlah <i>tree</i> dalam model
Subsample	Floating-point	Proporsi sampel data untuk setiap <i>tree</i>
Colsample_bytree	Floating-point	Proporsi fitur yang digunakan untuk

		membentuk setiap <i>tree</i>
Gamma	Floating-point	Regularisasi untuk mengontrol proses <i>pruning</i>
Min_child_weight	Floating-point	Mengontrol jumlah bobot untuk setiap <i>leaf</i> pada <i>tree</i>
Learning_rate	Floating-point	Proporsi penambahan residual untuk setiap pembelajaran
Max_depth	Integer	Kedalaman maksimum setiap <i>tree</i> pada model
Reg_alpha	Floating-point	Regularisasi melalui pemberian penalti

2.2.4. Pengujian dan Evaluasi Model

Setelah proses pelatihan dan optimasi selesai, langkah selanjutnya adalah melakukan pengujian dan evaluasi model guna menilai

kinerjanya terhadap data *testing* serta mengukur kemampuan generalisasinya dalam melakukan prediksi.



Gambar 4. Tahapan Pengujian dan Evaluasi Model

Gambar 4 menunjukkan proses pengujian dan evaluasi, di mana model hasil optimasi diuji menggunakan data *testing* dan dievaluasi dengan empat metrik *error*, yaitu RMSE, MAE, MFE, dan MAPE. Hasil evaluasi kemudian dibandingkan dengan model *default* untuk menilai efektivitas model yang diusulkan.

3. HASIL DAN PEMBAHASAN

3.1. Karakteristik Dataset

Pada penelitian ini, data diperoleh langsung dari Badan Pusat Statistik (BPS) Kabupaten Buleleng dan mencakup dua variabel utama, yaitu luas panen (hektar) serta produksi padi (ton). Produktivitas padi dihitung berdasarkan perbandingan antara total produksi dan luas panen, dengan satuan ton per hektar. Data dikumpulkan untuk periode Januari 2018 hingga Desember 2024, menghasilkan total 84 baris untuk masing-masing variabel produksi dan produktivitas. Dataset tersebut dapat diakses melalui situs resmi BPS Kabupaten Buleleng pada bagian Produksi Padi per Bulan Menurut Kabupaten/Kota di Provinsi Bali dan Luas Panen per Bulan Menurut Kabupaten/Kota di Provinsi Bali.

3.2. Pra-pemrosesan Data

Setelah proses pembersihan data (*data cleaning*) selesai, langkah selanjutnya adalah tahap pemilihan fitur (*feature selection*). Pada tahap ini, dilakukan analisis untuk mengidentifikasi variabel-variabel yang

memiliki pengaruh paling signifikan terhadap target, yaitu produksi dan produktivitas padi, dengan memanfaatkan metode seperti *Feature Importance* dan *Cross Correlation Function* (CCF).

Tabel 2. Jumlah Fitur untuk Masing-masing Dataset

Dataset	Metode	Time-step
Produksi Padi	Feature Importance	12
	CCF	12
Produktivitas Padi	Feature Importance	1

Berdasarkan hasil seleksi fitur menggunakan metode *Feature Importance* dan CCF, diperoleh bahwa dataset produksi padi memiliki 24 fitur signifikan, 12 fitur ditentukan melalui *Feature Importance* dan 12 fitur lainnya berasal dari variabel luas panen yang diidentifikasi menggunakan CCF. Sementara itu, untuk dataset produktivitas padi, hanya terdapat satu fitur relevan berdasarkan hasil *Feature Importance*, sebagaimana ditunjukkan pada tabel sebelumnya. Berdasarkan temuan tersebut, proses pengolahan data dilanjutkan secara bertahap, dimulai dengan mentransformasikan (*reframing*) data menjadi bentuk *cross-sectional* menggunakan metode *sliding window* sesuai jumlah *time-step* yang diperoleh dari tahap seleksi fitur.

Pembagian data dilakukan secara kronologis dengan rasio 8:2, di mana data hasil

sliding window untuk produksi padi mencakup periode Januari 2019 hingga Oktober 2023 sebagai data pelatihan, serta November 2023 hingga Desember 2024 sebagai data pengujian. Sementara itu, pada data produktivitas padi, periode Februari 2018 hingga April 2023 digunakan untuk pelatihan dan Mei 2023 hingga Desember 2024 digunakan untuk pengujian. Selanjutnya, dilakukan normalisasi data sebelum masuk ke tahap optimasi model.

3.3. Optimasi Model

3.3.1 Hiperparameter GA

Tabel 3 dan 4 menyajikan rincian konfigurasi hiperparameter yang digunakan untuk optimasi model XGBoost pada target produksi dan produktivitas padi, mencakup nama hiperparameter, tipe data, serta nilai yang digunakan dalam proses eksperimen.

Tabel 3. Rincian Hiperparameter GA untuk Produksi Padi

No	Hiperparameter	Nilai
1	Generasi	40
2	Jumlah induk (parent)	2
3	Populasi	20
4	Crossover rate	(0.8, 0.6)
5	Mutation rate	0.1
6	Indeks Distribusi (η)	30

Tabel 4. Rincian Hiperparameter GA untuk Produktivitas Padi

No	Hiperparameter	Nilai
1	Generasi	60
2	Jumlah induk (parent)	2
3	Populasi	20
4	Crossover rate	(0.8, 0.6)
5	Mutation rate	0.1
6	Indeks Distribusi (η)	30

3.3.2 Hiperparameter XGBoost

Selain melakukan inialisasi pada hiperparameter GA, nilai-nilai hiperparameter XGBoost juga perlu ditentukan terlebih dahulu pada tahap awal proses optimasi. Setelah itu, GA membentuk populasi awal secara acak, di mana setiap kromosom berisi sejumlah gen yang masing-masing merepresentasikan satu nilai hiperparameter dari XGBoost. Rincian hiperparameter yang dioptimasi disajikan pada Tabel 5.

Tabel 5. Hiperparameter XGBoost yang dioptimasi

No	Hiperparameter	Rentang Nilai
1	N estimator (ne)	(100, 1000)
2	Subsample (ss)	1
3	Colsample_bytree (csb)	(0.9, 1.0)
4	Gamma (gm)	(0.01, 0.1)
5	Min_child_weight (mcw)	(0.01, 10)
6	Learning rate (lr)	0.3
7	Max_depth (md)	6
8	Reg_alpha (ra)	(0.0, 0.2)

3.3.3 Tipe Operator GA

GA memiliki beberapa operator utama yang mendukung proses pencarian solusi optimal, yaitu operator seleksi, *crossover* (persilangan), dan mutasi. Tipe operator yang digunakan dalam penelitian ini disajikan secara rinci pada Tabel 6 berikut.

Tabel 6. Tipe Operator pada GA

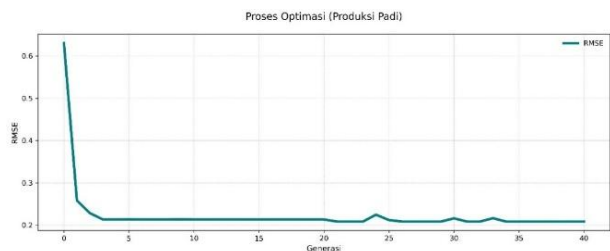
No	Operator	Tipe
1	Basis Seleksi	Nilai Fitness
2	Seleksi	Tournament Selection
3	Crossover	Uniform Crossover + Linear Adaptive
4	Mutasi	Polynomial Mutation

Pemilihan tipe pada masing-masing operator dilakukan dengan mempertimbangkan kesesuaian terhadap karakteristik masalah, kestabilan performa pada berbagai ukuran populasi, serta rekomendasi dari literatur yang relevan. Sebagai ilustrasi, penerapan *uniform crossover* yang dipadukan dengan *linear adaptive* bertujuan untuk menyeimbangkan eksplorasi dan eksploitasi dengan menurunkan *crossover rate* secara bertahap. Jika merujuk pada Tabel 3 dan 4, nilai *crossover* diatur dari 0.8 pada awal evolusi menjadi 0.6 pada tahap akhir.

3.3.4 Iterasi dan Penghentian Proses

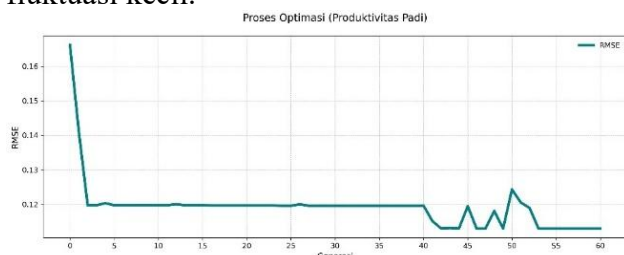
Proses iterasi berperan penting dalam mengarahkan pencarian solusi optimal dalam proses optimasi. Selama iterasi berlangsung, nilai *fitness* model dievaluasi dengan metrik RMSE untuk menilai kualitas solusi yang dihasilkan. Proses optimasi dilakukan hingga mencapai jumlah generasi maksimum atau hingga nilai RMSE mengalami konvergensi.

Grafik pada gambar 5 dan 6 menunjukkan nilai RMSE dari individu terbaik pada tiap generasi untuk data produksi dan produktivitas padi.



Gambar 5. Optimasi Model untuk Produksi Padi

Gambar 7 menunjukkan nilai RMSE mengalami tren penurunan seiring bertambahnya generasi. Berdasarkan gambar di atas, setelah generasi ke-4 nilai RMSE mulai stabil di angka 0.3 ke bawah dan mengalami fluktuasi kecil.



Gambar 6. Optimasi Model untuk Produktivitas Padi

Sama seperti produksi padi, grafik RMSE produktivitas padi yang ditunjukkan pada Gambar 8, menunjukkan penurunan signifikan hingga generasi ke-4, kemudian stabil dengan fluktuasi kecil di bawah 0.12 hingga generasi ke-40. Pada generasi ke-41 hingga 50, nilai RMSE berfluktuasi di kisaran 0.117 hingga 0.125 dan mulai generasi ke-52 hingga ke-60, nilai RMSE kembali stabil di sekitaran 0.117.

3.4. Pengujian dan Evaluasi Model

3.4.1. Hiperparameter Model

Tabel 7 dan 8 menampilkan perbandingan kombinasi *hyperparameter* antara model XGBoost yang dioptimalkan menggunakan GA dan model XGBoost tanpa proses optimasi pada data produksi serta produktivitas padi. Analisis terhadap perbedaan nilai *hyperparameter* dilakukan untuk mengetahui bagaimana konfigurasi masing-masing model berpengaruh terhadap kinerja selama tahap pelatihan dan pengujian.

Tabel 7. Perbandingan Hiperparameter Model untuk Produksi Padi

No	Hiperparameter	GA-XGBoost	XGBoost
1	n_estimator	262	100
2	subsample	1.0	1.0
3	colsample_bytree	0.97	1.0
4	gamma	0.1	0.0
5	min_child_weight	0.3005	1.0
6	learning_rate	0.3	0.3
7	max_depth	6	6
8	reg_alpha	0.151558	0.0

Tabel 8. Perbandingan Hiperparameter Model untuk Produktivitas Padi

No	Hiperparameter	GA-XGBoost	XGBoost
1	n_estimator	246	100
2	subsample	1.0	1.0
3	colsample_bytree	0.99	1.0
4	gamma	0.029911	0.0
5	min_child_weight	0.81922	1.0
6	learning_rate	0.3	0.3
7	max_depth	6	6
8	reg_alpha	0.01172	0.0

3.4.2. Evaluasi Model

A. Produksi Padi

Evaluasi model dilakukan untuk mengukur performa model terhadap data latih dan data uji. Tabel 9 berikut menyajikan hasil evaluasi terhadap model prediksi produksi padi berdasarkan keempat matriks evaluasi yang digunakan.

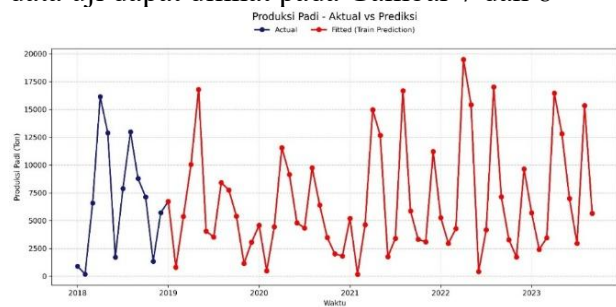
Tabel 9. Hasil Evaluasi Model untuk Data Produksi Padi

Model	Metrik Evaluasi	Train	Test
-------	-----------------	-------	------

GA-XGBoost	RMSE	0.208684	2,391.147
	MAE	0.152739	1,906.015
	MFE	-	923.306
	MAPE (%)	0.010494	25.489

Berdasarkan Tabel 9. hasil pengujian menunjukkan bahwa model GA-XGBoost menghasilkan nilai RMSE sebesar 2,391.147, MAE 1,906.015, MFE 923.306, dan MAPE 25,5%. Sementara itu, pada data latih, nilai kesalahannya jauh lebih kecil, yakni RMSE 0.2087, MAE 0.1527, dan MAPE 0.0105%. Perbedaan yang cukup besar antara performa pada data latih dan data uji ini mengindikasikan adanya *overfitting*, yaitu kondisi ketika model terlalu menyesuaikan diri terhadap pola pada data pelatihan sehingga kemampuannya menggeneralisasi data baru menurun.

Setelah mengevaluasi performa model pada kedua dataset tersebut, langkah selanjutnya adalah melakukan visualisasi hasil prediksi. Hasil visualisasi model terhadap data latih dan data uji dapat dilihat pada Gambar 7 dan 8



Gambar 7. Prediksi GA-XGBoost terhadap Data Pelatihan Produksi Padi



Gambar 8. Prediksi GA-XGBoost terhadap Data Uji Produktivitas Padi

B. Produktivitas Padi

Evaluasi terhadap kinerja model GA-XGBoost dilakukan untuk mengukur sejauh mana model mampu memprediksi data

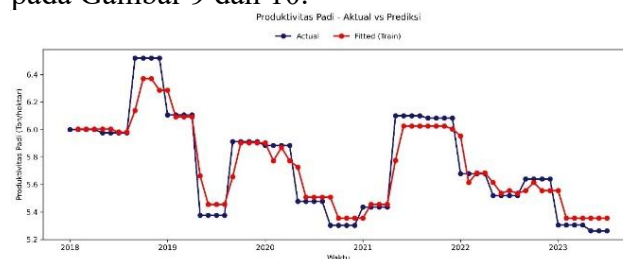
produktivitas padi dengan baik. Pengujian ini menggunakan metrik RMSE, MAE, MFE, dan MAPE pada data latih dan data uji untuk menilai tingkat generalisasi model. Hasil ringkasan evaluasi tersebut disajikan pada Tabel 10.

Tabel 10. Hasil Evaluasi Model untuk Data Produktivitas Padi

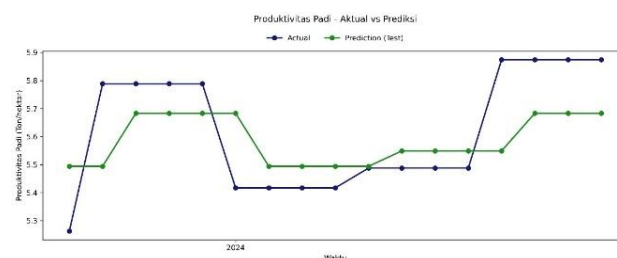
Model	Metrik Evaluasi	Train	Test
GA-XGBoost	RMSE	0.119564	0.180151
	MAE	0.081684	0.132577
	MFE	-0.00017	-0.06889
	MAPE (%)	1.416517	2.317327

Berdasarkan Tabel 10, hasil pengujian menunjukkan bahwa peningkatan nilai *error* antara data pelatihan dan pengujian relatif kecil, dengan selisih RMSE sekitar 0.061, MAE sebesar 0.051, dan MAPE sebesar 0.9%. Nilai MFE negatif sebesar -0.0699 mengindikasikan bahwa model cenderung sedikit *overestimate*.

Berdasarkan hasil evaluasi tersebut, kondisi ini disebabkan oleh distribusi nilai produktivitas padi pada data uji yang lebih rendah dibandingkan dengan rata-rata historis yang dipelajari model selama proses pelatihan. Setelah evaluasi kinerja model pada data latih dan uji dilakukan, tahap berikutnya adalah visualisasi hasil prediksi. Visualisasi prediksi model untuk kedua jenis dataset ditampilkan pada Gambar 9 dan 10.



Gambar 9. Prediksi GA-XGBoost terhadap Data Latih Produktivitas Padi



Gambar 10. Prediksi GA-XGBoost terhadap Data Uji Produktivitas Padi

3.5. Perbandingan Model

3.5.1. Metrik Evaluasi

A. Produksi Padi

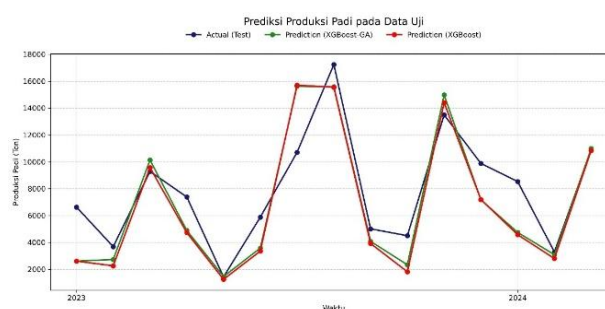
Tabel 11 menunjukkan perbandingan performa model GA-XGBoost dan XGBoost

Tabel 11. Perbandingan Performa Model untuk Data Produksi Padi

Model	Proses	RMSE	MAE	MAPE
GA-XGBoost	Train	0.208684	0.152739	0.010494
	Test	2,391.14	1,906.015	25.489
XGBoost	Train	0.361345	0.193418	0.018258
	Test	2,476.35	1,976.73	28.475

Pada data pengujian, model GA-XGBoost tetap menunjukkan kinerja yang lebih baik dengan nilai RMSE sebesar 2,391.15, MAE sebesar 1,906.015, dan MAPE sebesar 25.49%, dibandingkan dengan model XGBoost yang memiliki nilai *error* lebih tinggi, yaitu RMSE 2,476.353, MAE 1,976.728, dan MAPE 28.475%. Meskipun terjadi peningkatan *error* pada kedua model, perbedaan nilai *error* tersebut menunjukkan bahwa penerapan GA mampu meningkatkan kemampuan generalisasi model, sehingga tingkat kesalahan dapat ditekan meskipun peningkatannya tidak terlalu besar. Untuk melihat kemampuan model dalam memprediksi data uji, hasil perbandingan prediksi XGBoost dan GA-XGBoost terhadap data aktual disajikan pada Gambar 11.

dalam memprediksi produksi padi. GA-XGBoost secara konsisten memberikan hasil yang lebih baik dibandingkan XGBoost tanpa optimasi, baik pada tahap pelatihan maupun pengujian.



Gambar 11. Prediksi Model terhadap Data Uji Produktivitas Padi

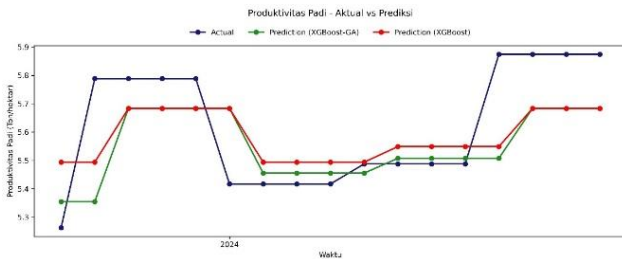
B. Produktivitas Padi

Tabel 12 menunjukkan perbandingan kinerja model GA-XGBoost dan XGBoost dalam memprediksi data latih dan uji produktivitas padi. Secara umum, GA-XGBoost menghasilkan nilai *error* lebih rendah pada tahap pelatihan, dengan RMSE 0.1196, MAE 0.0817, dan MAPE 1.42%, dibandingkan XGBoost (RMSE 0.208, MAE 0.1419, dan MAPE 2.46%).

Tabel 12. Perbandingan Performa Model untuk Data Produktivitas Padi

Model	Proses	RMSE	MAE	MAPE
GA-XGBoost	Train	0.11956	0.081684	1.41651
	Test	0.18015	0.132577	2.31732
XGBoost	Train	0.20803	0.141926	2.46080
	Test	0.16968	0.142825	2.52264

Pada data uji, GA-XGBoost memperoleh MAPE sebesar 2,31% dan MAE 0.1325 lebih rendah dibanding XGBoost, namun mencatatkan kuadrat kesalahan yang justru lebih tinggi. Untuk menganalisis lebih lanjut terkait hal tersebut, performa prediksi kedua model ditampilkan pada Gambar 12.



Gambar 12. Prediksi Model terhadap Data Uji Produktivitas Padi.

Tabel 13 di bawah, menyajikan perbandingan nilai aktual dengan hasil prediksi dari kedua model beserta nilai *error* masing-masing.

Tabel 13. Perbandingan Nilai Error untuk Data Uji Produktivitas Padi

No	Nilai	GA-XGBoost		XGBoost	
		Abs. Error	Sqr. Error	Abs. Error	Sqr. Error
1	5.26248	0.09233	0.00852	0.23149	0.05359
2	5.78869	0.43388	0.18825	0.29471	0.08685
3	5.78870	0.10483	0.01099	0.10513	0.01105
4	5.78870	0.10483	0.01099	0.10513	0.01105
5	5.78870	0.10483	0.01099	0.10513	0.01105
6	5.41714	0.26673	0.07115	0.26642	0.07098
7	5.41715	0.03808	0.00145	0.07681	0.00590
8	5.41715	0.03808	0.00145	0.07681	0.00590
9	5.41715	0.03808	0.00145	0.07681	0.00590
10	5.48776	0.03253	0.00106	0.00621	0.00004
11	5.48776	0.01984	0.00039	0.06122	0.00375
12	5.48776	0.01984	0.00039	0.06122	0.00375
13	5.48776	0.01984	0.00039	0.06122	0.00375
14	5.87480	0.36720	0.13484	0.32582	0.10616
15	5.87483	0.19096	0.03647	0.19126	0.03658
16	5.87487	0.19100	0.03648	0.19130	0.03660
17	5.87482	0.19095	0.03646	0.19125	0.03657
Rata-rata		0.13257	0.03245	0.14282	0.02879

Berdasarkan nilai *error* absolut (*abs. error*) dan *error* kuadrat (*sqr. error*) pada tabel di atas, nilai RMSE pada GA-XGBoost (0.1802) yang sedikit lebih tinggi dibandingkan XGBoost (0.1697) dapat dijelaskan secara teknis dengan mempertimbangkan distribusi *error* yang tercatat pada masing-masing titik. Secara khusus, nilai *error* kuadrat terbesar pada GA-XGBoost tercatat pada titik ke-2 dengan nilai *abs. error* sebesar 0.188255, diikuti oleh titik ke-14 sebesar 0.13484, titik ke-6 sebesar 0.07115, dan titik ke-10 sebesar 0.00106. Keempat nilai berkontribusi signifikan terhadap peningkatan nilai RMSE, hal ini karena RMSE cenderung mengkuadratkan nilai *error* sehingga sensitif terhadap *outlier*.

Namun, jika ditinjau dari jumlah prediksi dengan *sqr. error* yang lebih kecil, GA-XGBoost unggul 14 dari 17 titik data secara keseluruhan, yang menunjukkan kestabilan model. Hal ini juga tercermin dari nilai MAE dan

MAPE yang lebih rendah, menunjukkan GA-XGBoost secara umum lebih akurat terhadap nilai aktual.

3.5.2. Waktu Eksekusi

A. Produksi Padi

Evaluasi terkait waktu eksekusi dilakukan untuk mengukur efisiensi komputasi dari masing-masing model dalam proses pelatihan dan prediksi. Tabel 14 berikut menunjukkan rata-rata waktu pelatihan dan prediksi (*inference*) dari 10 percobaan yang dilakukan.

Tabel 14. Waktu Pelatihan dan Prediksi untuk Data Produksi Padi

Model	Waktu (detik)	
	Pelatihan	Prediksi
XGBoost	0.07171	0.00113
GA-XGBoost	0.08143	0.00264

Berdasarkan Tabel 14, perbedaan durasi pelatihan antara GA-XGBoost dan XGBoost dipengaruhi oleh variasi konfigurasi hiperparameter pada masing-masing model. GA-XGBoost menggunakan 262 *n_estimators*, jumlah yang lebih besar dibandingkan XGBoost yang hanya menggunakan 100 *n_estimators*. Secara umum, peningkatan jumlah *n_estimators* berdampak pada lamanya proses *boosting* yang dilakukan. Selain itu, nilai *min_child_weight* pada GA-XGBoost yang lebih kecil memungkinkan terbentuknya lebih banyak *leaf* pada pohon keputusan, yang pada akhirnya turut memperpanjang waktu pelatihan serta proses prediksi model.

B. Produktivitas Padi

Seperti halnya pada model produksi padi, jumlah *n_estimators* pada GA-XGBoost yaitu sebanyak 246 *tree*, mempengaruhi lama waktu *boosting*.

Tabel 15. Waktu Pelatihan dan Prediksi untuk Data Produktivitas Padi

Model	Waktu (detik)	
	Pelatihan	Prediksi
XGBoost	0.02866	0.0023

Tabel 16. Perbandingan Model dalam Prediksi Produksi Padi

Algoritma	RMSE		MAE		MAPE (%)	
	Train	Test	Train	Test	Train	Test
Decision Tree	525	4,760	336	3,679	525	4,760
GBM	95.50	3,049	79.63	2,552	95.50	3,049
GA-XGBoost	0.2086	2,391	0.1527	1,906	0.2086	2,391
RF-Regressor	941	2,704	734	2,276	941	2,704
MLP	1,965	3,270	1,306	2,686	1,965	3,270

Berdasarkan evaluasi pada Tabel 16, GA-XGBoost memperoleh nilai RMSE, MAE, dan MAPE terendah pada data pelatihan maupun pengujian untuk prediksi produksi padi, menunjukkan kemampuan generalisasi yang lebih baik relatif terhadap model lain. GBM dan RF-Regressor memberikan kinerja cukup baik, tetapi masih berada di bawah GA-XGBoost.

Tabel 17. Perbandingan Model untuk Prediksi Produktivitas Padi.

Algoritma	RMSE		MAE		MAPE (%)	
	Train	Test	Train	Test	Train	Test
Decision Tree	0.2197	0.2406	0.1602	0.1923	0.2197	0.2406
GBM	0.201	0.178	0.1676	0.1607	0.201	0.178

GA-XGBoost	0.09928	0.00316
------------	---------	---------

Waktu prediksi GA-XGBoost, yaitu sekitar 0.00316 detik per prediksi, tercatat lebih lama dibandingkan XGBoost yang memerlukan sekitar 0.0023 detik. Perbedaan ini konsisten dengan hasil waktu pelatihan, di mana faktor utama penyebabnya adalah jumlah *n_estimators* yang lebih besar, sehingga setiap sampel data perlu melalui lebih banyak *tree* untuk menghasilkan prediksi. Selain itu, nilai *min_child_weight* yang lebih kecil juga memungkinkan struktur *tree* menjadi lebih dalam, yang pada gilirannya memperpanjang proses traversal selama tahap prediksi.

3.5.3. Machine Learning

A. Produksi Padi

Dengan membandingkan beberapa model lain, kita dapat menilai secara jelas bagaimana perbedaan struktur dan karakter algoritma memengaruhi prediksi pada data latih dan uji, sembari menempatkan GA-XGBoost sebagai model utama.

B. Produktivitas Padi

Berdasarkan Tabel 17, MLP menjadi model yang paling konsisten pada data latih maupun uji. MLP memperoleh nilai MAE, RMSE, dan MAE terendah, menunjukkan kestabilan kinerja dibanding model lain.

GA-XGBoost	0.1195	0.1801	0.0816	0.1325	0.1195	0.1801
RF Regressor	0.07	0.20	0.05	0.11	0.07	0.20
MLP	0.11	0.123	0.073	0.08	1.27	1.42

Stabilitas tersebut dicapai melalui arsitektur MLP, yang terdiri atas dua hidden layer (8 dan 4 neuron) dengan aktivasi LeakyReLU. Hiperparameter seperti *learning rate*, jumlah *epoch*, dan *batch size* disetel sesuai karakteristik jumlah fitur pada data produktivitas sehingga pembaruan bobot dan bias berjalan optimal.

4. PENUTUP

4.1. Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah disajikan pada bab sebelumnya, dapat disimpulkan bahwa optimasi model XGBoost menggunakan *Genetic Algorithm* (GA) dalam penelitian ini dirancang secara sistematis dan adaptif untuk meningkatkan performa prediksi. Proses optimasi diawali dengan inisialisasi hiperparameter GA, penentuan rentang nilai hiperparameter XGBoost, serta pemilihan operator yang sesuai dengan *tournament selection* untuk seleksi, *linear adaptive uniform crossover* untuk penyilangan, dan *polynomial mutation* untuk mutasi. Melalui pendekatan ini, diperoleh kombinasi hiperparameter optimal yang mampu meningkatkan performa XGBoost, meskipun tidak secara signifikan, namun memberikan kemampuan generalisasi yang lebih baik pada data uji dibandingkan model tanpa optimasi. Evaluasi menggunakan metrik RMSE, MAE, dan MAPE menunjukkan hasil yang secara konsisten lebih rendah pada tahap pelatihan maupun pengujian. Pada data uji, GA–XGBoost menurunkan galat prediksi produksi padi dengan RMSE sebesar 2,391 ton, MAE 1,906 ton, dan MAPE 25.49%; sedangkan pada XGBoost tanpa optimasi masing-masing sebesar 2,476 ton, 1,977 ton, dan 28.48%. Untuk produktivitas, GA–XGBoost mencapai MAE 0.133 ton/ha dan MAPE 2.32%, sedangkan XGBoost tanpa optimasi 0.143 ton/ha dan 2.52%; namun RMSE GA–XGBoost tercatat

0.180 ton/ha, lebih tinggi dibanding 0,170 ton/ha pada XGBoost

Namun demikian, performa GA-XGBoost pada prediksi produksi padi menunjukkan adanya perbedaan signifikan antara hasil pelatihan dan pengujian, yang mengindikasikan kecenderungan *overfitting* akibat dua faktor utama yaitu keterbatasan jumlah data yang mengurangi keragaman pola yang dapat dipelajari model, serta konfigurasi hiperparameter yang belum sepenuhnya selaras dengan karakteristik data yang digunakan.

4.2. Saran

Berdasarkan hasil penelitian ini, terdapat beberapa rekomendasi yang dapat dijadikan pertimbangan untuk pengembangan penelitian selanjutnya, sebagaimana dijelaskan berikut.

A. Penambahan Data

Penambahan dataset yang lebih beragam dapat memperluas kemampuan prediksi model, terutama untuk model berbasis *tree* seperti XGBoost. Dengan lebih banyak data, pola lokal dalam data aktual dapat dikenali secara lebih general. Selain itu, peningkatan jumlah data memungkinkan proses pemodelan menjadi lebih adaptif, sehingga pengaturan hiperparameter standar dari model tidak perlu dikurangi.

B. Penyesuaian Rentang Nilai Hiperparameter

Penentuan rentang nilai *learning_rate*, *max_depth*, *gamma*, *min_child_weight*, *subsample*, atau hiperparameter lainnya pada XGBoost sebaiknya disesuaikan dengan karakteristik data, terutama jika jumlah data terbatas atau distribusinya relatif sempit.

C. Penambahan Algoritma Metaheuristik Lain

Penggunaan algoritma metaheuristik tambahan dapat dipertimbangkan untuk memperluas ruang pencarian solusi. Sebagai contoh, integrasi algoritma *Cuckoo Search* (CS) ke dalam proses *crossover* dan mutasi pada *Genetic Algorithm* (GA) dapat meningkatkan keberagaman solusi sekaligus mempercepat

konvergensi. Selain itu, penerapan *Grey Wolf Optimizer* (GWO) pada tahap *crossover* GA berpotensi mempercepat konvergensi dan membantu model menemukan solusi yang lebih optimal.

D. Integrasi dengan Model Statistik

Model statistik, seperti *Autoregressive* (AR), dapat dimanfaatkan untuk menentukan lag yang signifikan pada data *time-series* univariat, sedangkan VAR cocok digunakan untuk data multivariat. Lag-lag yang diidentifikasi ini kemudian dapat dijadikan sebagai fitur input dalam model *machine learning*.

E. Penambahan Variabel Eksogen

Penambahan variabel eksogen seperti curah hujan (mm/bulan), suhu (°C), kelembapan, serta indeks vegetasi (mis. NDVI/EVI) dapat menangkap pengaruh musim dan kondisi agroklimat terhadap produksi/produktivitas.

5. DAFTAR PUSTAKA

- Rahayu, P.I.R., Hidayatullah, M. and Hijrah, M. (2023) "Implementation Vector Autoregressive (Var) On Rice Production and Rice Productivity Data in Indonesia," *Jurnal Matematika, Statistika dan Komputasi*, 19(3), pp. 580–592. <https://doi.org/10.20956/j.v19i3.24881>.
- Tabassum, R. et al. (2024) "Machine learning-based prediction of scour depth evolution around spur dikes," *Journal of Hydroinformatics*, 26(11), pp. 2815–2836. <https://doi.org/10.2166/hydro.2024.160>.
- Swain, K.C., Singha, C. and Pradhan, B. (2024) "Estimating Total Rice Biomass and Crop Yield at Field Scale Using PlanetScope Imagery Through Hybrid Machine Learning Models," *Earth Systems and Environment*, 8(4), pp. 1713–1731. <https://doi.org/10.1007/s41748-024-00481-2>.
- Noorunnahar, M., Chowdhury, A.H. and Mila, F.A. (2023) "A tree based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh," *PLoS ONE*, 18(3 March). <https://doi.org/10.1371/journal.pone.0283452>.
- Dewi, N.P.N.P. et al. (2025) "Enhancing Diesel Backup Power Forecasting With LSTM, GRU, and Autoencoder-based Input Encoding," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 14(1), pp. 183–199. <https://doi.org/10.23887/janapati.v14i1.92079>.
- Handayani Putri, D.P. et al. (2023) "Perbandingan Performansi Support Vector Machine (Svm) dan Backpropagation untuk Klasifikasi Studi Mahasiswa Undiksha," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 9(3), p. 492. <https://doi.org/10.26418/jp.v9i3.67843>.
- Pascima, I.B.N. and Putrama, I.M. (2021) "Forecasting foreign exchange rate using a combination of linear regression and flower pollination algorithm," in *Journal of Physics: Conference Series*. IOP Publishing Ltd. <https://doi.org/10.1088/1742-6596/1810/1/012021>.
- Putu Novita Puspa Dewi, N. et al. (2021) "OPTIMASI GENERAL REGRESSION NEURAL NETWORK DENGAN FRUIT FLY OPTIMIZATION ALGORITHM UNTUK PREDIKSI PEMAKAIAN ARUS LISTRIK PADA PENYULANG," 18(1), pp. 1–12. <https://journal.unpak.ac.id/index.php/komputasi>
- Dewi, N.P.N.P., Kertiasih, N.K. and Sintiar, N.L.D. (2022) "Modifikasi Fruit Fly Optimization Algorithm untuk Optimasi General Regression Neural Network pada Kasus Prediksi Time-Series," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 11(3), pp. 192–204. <https://doi.org/10.23887/janapati.v11i3.54521>.
- Tran, T.N. and Nguyen, Q.D. (2024) "Research on the Influence of Genetic Algorithm Parameters on XGBoost in Load Forecasting," *Engineering, Technology and Applied Science Research*, 14(6), pp. 18849–18854. <https://doi.org/10.48084/etasr.8863>.
- Zhang, R. and Jia, H. (2021) "Production performance forecasting method based on multivariate time series and vector autoregressive machine learning model for waterflooding reservoirs," *Petroleum Exploration and Development*, 48(1), pp. 201–211. [https://doi.org/10.1016/S1876-3804\(21\)60016-2](https://doi.org/10.1016/S1876-3804(21)60016-2).

- Jiang, P. et al. (2023) “Prediction Dynamics in Cotton Aphid Using Unmanned Aerial Vehicle Multispectral Images and Vegetation Indices,” *IEEE Access*, 11, pp. 5908–5918. Available at: <https://doi.org/10.1109/ACCESS.2023.3235912>
- Aydin, M.R. and Gülsün, B. (2024) “Optimizing a Machine Learning Algorithm by Novel Metaheuristic Approach: Case Study in Forecasting.” Available at: <https://doi.org/10.20944/preprints202411.0907.v1>.
- Dedy, I.P. et al. (2023) “Optimasi Parameter Support Vector Machine Dengan Algoritma Genetika Untuk Analisis Sentimen Pada Media Sosial Instagram.” Available at: <https://doi.org/10.31598>.
- Samantaray, S., Sahoo, A. and Baliarsingh, F. (2024) “Groundwater level prediction using an improved SVR model integrated with hybrid particle swarm optimization and firefly algorithm,” *Cleaner Water*, 1. Available at: <https://doi.org/10.1016/j.clwat.2024.100003>
- Fang, Z.G. et al. (2022) “Application of a data-driven XGBoost model for the prediction of COVID-19 in the USA: a time-series study,” *BMJ Open*, 12(7). Available at: <https://doi.org/10.1136/bmjopen-2021-056685>.
- Rathod, S. et al. (2021) “Two-stage spatiotemporal time series modelling approach for rice yield prediction & advanced agroecosystem management,” *Agronomy*, 11(12). Available at: <https://doi.org/10.3390/agronomy11122502>.
- Oukhouya, H. et al. (2024) “Forecasting International Stock Market Trends: XGBoost, LSTM, LSTM-XGBoost, And Backtesting XGBoost Models,” *Statistics, Optimization and Information Computing*, 12(1), pp. 200–209. Available at: <https://doi.org/10.19139/soic-2310-5070-1822>.
- Li, J. et al. (2022) Application of XGBoost Algorithm in the Optimization of 2 Pollutant Concentration. Available at: <https://ssrn.com/abstract=4029398>.
- Sriartha, I.P., Giyarsih, S.R. and Purnamawati, I.G.A. (2023) “Comparing the adaptive capacity of traditional irrigated rice fields farmers in urban and rural areas to climate change in Bali, Indonesia,” *Cogent Social Sciences*, 9(2). <https://doi.org/10.1080/23311886.2023.2275936>.