

Pengembangan Model Prediksi Produksi Padi di Provinsi Jawa Barat Berbasis Integrasi Data Agroklimat Menggunakan *Ensemble Machine Learning*

Development of a Rice Production Prediction Model in West Java Based on Integrated Agroclimatic Data Using Ensemble Machine Learning

Abil Fadila^{1*}, Muhammad Farrel Syabena², Miftakhul Bakhrrir Rozaq Khamid¹, Hadi Prastyo¹, Sarah Sakinah Umadi¹, Ilmiasa Saliha¹

¹Departemen Agroteknologi, Fakultas Pertanian, Universitas Singaperbangsa Karawang, Karawang 41361, Indonesia

²Departemen Agribisnis, Fakultas Pertanian, Universitas Singaperbangsa Karawang, Karawang 41361, Indonesia

*E-mail: abil.fadila@faperta.unsika.ac.id

Diterima: 6 April 2026; Disetujui: 6 Agustus 2026

ABSTRAK

Dinamika iklim yang tidak menentu menjadi faktor utama yang mempengaruhi variabilitas produksi padi di Jawa Barat dan menuntut penerapan teknologi prediksi yang akurat untuk mendukung pengelolaan produksi pertanian secara efisien dan adaptif. Penelitian ini bertujuan mengembangkan sistem prediksi produksi padi berbasis integrasi data agroklimat menggunakan algoritma *ensemble machine learning*, yaitu *Random Forest* (RF) dan *Extreme Gradient Boosting* (XGBoost). Data yang digunakan meliputi variabel agroklimat berupa curah hujan, suhu udara, kelembapan, dan kecepatan angin serta data historis produksi padi di Provinsi Jawa Barat periode 2018–2024. Pengembangan sistem dilakukan melalui tahapan pra-pemrosesan data, transformasi fitur, pelatihan model, dan validasi menggunakan skema *10-fold cross-validation* untuk menjamin stabilitas dan reliabilitas model. Kinerja model dievaluasi menggunakan koefisien determinasi (R^2), *Mean Absolute Error* (MAE), dan *Root Mean Square Error* (RMSE). Hasil penelitian menunjukkan bahwa algoritma *Random Forest* memberikan performa terbaik dengan nilai R^2 sebesar 0,9928, MAE sebesar 16.608,92, dan RMSE sebesar 24.796,55. Variabel luas lahan, curah hujan, suhu, dan kelembapan teridentifikasi sebagai faktor utama yang memengaruhi produksi padi. Sistem yang dikembangkan berpotensi diimplementasikan sebagai teknologi pendukung prediksi produksi padi dengan memanfaatkan data agroklimat sebagai variabel prediktor, sehingga dapat mendukung pengambilan keputusan dalam pengelolaan produksi tanaman dan penerapan pertanian presisi berbasis data guna meningkatkan efisiensi serta ketahanan sistem produksi padi.

Kata kunci: *Ensemble machine learning*; pertanian presisi; pemodelan agroklimat; prediksi produksi padi; teknologi pertanian

ABSTRACT

Climate dynamics that are increasingly unpredictable have become a major factor affecting rice production variability in West Java, necessitating the implementation of accurate predictive technologies to support efficient and adaptive agricultural production management. This study aims to develop a rice production prediction system based on agroclimatic data integration using ensemble machine learning algorithms, namely *Random Forest* (RF) and *Extreme Gradient Boosting* (XGBoost). The data used include agroclimatic variables such as rainfall, air temperature, humidity, and wind speed, along with historical rice production data in West Java Province for the 2018–2024 period. System development was conducted through data preprocessing, feature transformation, model training, and validation using a *10-fold cross-validation* scheme to ensure model stability and reliability. Model performance was evaluated using the coefficient of determination (R^2), *Mean Absolute Error* (MAE), and *Root Mean Square Error* (RMSE). The results indicate that the *Random Forest* algorithm achieved the best performance with an R^2 value of 0.9928, MAE of 16,608.92, and RMSE of 24,796.55. Land area, rainfall, temperature, and humidity were identified as the main factors influencing rice production. The developed system has the potential to serve as a decision-support technology for predicting rice production by utilizing agroclimatic data as predictor variables. This system can support decision-making in crop production management and the implementation of data-driven precision agriculture, thereby enhancing the efficiency and resilience of rice production systems.

Keywords: *Agroclimate modeling*; crop yield prediction; ensemble machine learning; precision agriculture; rice production

PENDAHULUAN

Ketahanan pangan menjadi salah satu persoalan utama dalam pembangunan global yang memerlukan perhatian serius penerapan inovasi teknologi pada sistem produksi pertanian. Pangan, khususnya beras, merupakan komoditas strategis di Indonesia yang berperan penting dalam menjaga stabilitas ekonomi, sosial, dan ketahanan pangan nasional. Sebagai salah satu dari empat negara produsen beras terbesar di dunia, Indonesia menghadapi

tantangan dalam menjaga stabilitas produksi di tengah peningkatan permintaan dan dinamika perubahan iklim (Kharisma S et al., 2025). Menurut *Food and Agriculture Organization* (2021), padi merupakan komoditas pangan utama dengan kontribusi produksi global terbesar ketiga setelah jagung dan gandum, dengan dominasi produksi di kawasan Asia yang mencapai sekitar 90% dari total produksi dunia (Pratiwi et al., 2018). Dalam konteks nasional, Jawa Barat memiliki peran strategis sebagai salah satu sentra

produksi padi utama yang berkontribusi signifikan terhadap pasokan beras nasional.

Produksi padi di Jawa Barat menunjukkan dinamika yang tinggi dan cenderung fluktuatif. Data Badan Pusat Statistik Provinsi Jawa Barat menunjukkan bahwa produksi padi pada tahun 2024 mencapai 8.626.879,91 ton, namun mengalami variasi signifikan dalam beberapa tahun terakhir (BPS Jabar, 2024). Produksi padi sangat dipengaruhi oleh faktor agroklimat seperti curah hujan, suhu, dan kelembapan yang berpengaruh terhadap produktivitas lahan dan stabilitas hasil panen (Gunawan et al., 2024). Ketidakpastian kondisi agroklimat menuntut penerapan teknologi prediksi yang mampu memodelkan hubungan kompleks antara faktor lingkungan dan hasil produksi secara akurat guna mendukung pengelolaan produksi pertanian yang adaptif dan efisien.

Perkembangan teknologi komputasi dan *machine learning* membuka peluang dalam pengembangan sistem prediksi produksi pertanian berbasis data. Metode pembelajaran mesin mampu menangani data non linier dan multivariat yang kompleks serta meningkatkan akurasi prediksi dibandingkan pendekatan statistik konvensional (Sulardi et al., 2017). Beberapa penelitian sebelumnya menggunakan metode regresi linear untuk memprediksi hasil pertanian, termasuk pada komoditas hortikultura (Basahona et al., 2019) dan produksi padi (Ariyani & Arifin, 2021). Namun, pendekatan tersebut memiliki keterbatasan dalam merepresentasikan interaksi kompleks antar variabel agroklimat. Penelitian terbaru menunjukkan bahwa algoritma *ensemble machine learning* seperti *Random Forest* dan *Extreme Gradient Boosting (XGBoost)* mampu memberikan performa prediksi yang lebih robust pada data pertanian yang bersifat dinamis (Yasin et al., 2025). Meskipun demikian, sebagian besar penelitian masih berfokus pada variabel produksi konvensional seperti luas lahan dan produktivitas tanpa mengintegrasikan variabel agroklimat secara komprehensif pada skala regional (Hutahaean et al., 2024).

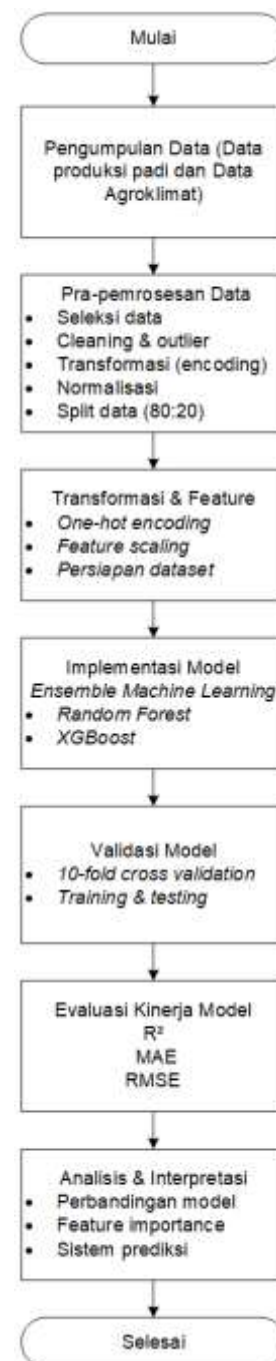
Meskipun berbagai penelitian telah mengembangkan model prediksi produksi pertanian menggunakan pendekatan statistik dan *machine learning*, sebagian besar studi masih menggunakan pendekatan statistik konvensional seperti regresi linear yang memiliki keterbatasan dalam prediksi hasil panen padi yaitu nilai akurasinya terbatas dalam merepresentasikan dinamika agroklimat yang kompleks (Herwanto et al., 2019). Studi yang menerapkan algoritma *machine learning* pada prediksi produksi padi umumnya berfokus pada variabel produksi konvensional seperti luas lahan dan produktivitas tanpa mengintegrasikan variabel agroklimat secara komprehensif pada skala regional. Sebagian besar penelitian juga masih bersifat analisis prediktif tanpa mengembangkan kerangka sistem prediksi yang terintegrasi sebagai teknologi pendukung pengelolaan produksi pertanian. Implementasi pendekatan *ensemble machine learning* untuk integrasi data agroklimat multivariat pada sistem produksi padi di wilayah tropis, khususnya pada skala regional, masih terbatas.

Berdasarkan keterbatasan tersebut, diperlukan pengembangan sistem prediksi produksi padi berbasis integrasi data agroklimat multivariat yang mampu memodelkan interaksi kompleks faktor lingkungan secara lebih akurat serta mendukung implementasi pertanian presisi. Oleh karena itu, kebaruan penelitian ini terletak pada pengembangan sistem prediksi produksi padi berbasis integrasi data agroklimat regional melalui perbandingan dua algoritma *ensemble learning*, yaitu *Random Forest* dan *Extreme Gradient Boosting (XGBoost)*, yang dipilih karena mewakili dua pendekatan *ensemble* yang berbeda, yaitu *bagging* dan *boosting*, serta telah banyak dilaporkan memiliki

kinerja yang baik pada permasalahan regresi non linier (Chen & Guestrin, 2016). Perbandingan kedua algoritma tersebut bertujuan untuk mengidentifikasi model yang memberikan performa prediksi terbaik pada data produksi padi berbasis agroklimat. Penelitian ini diharapkan menghasilkan model prediktif yang lebih adaptif terhadap kondisi lingkungan lokal serta mendukung pengambilan keputusan dalam pengelolaan produksi tanaman berbasis data.

METODOLOGI

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimental melalui pengembangan model prediktif berbasis *Machine Learning*. Tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Tabel 1. Identifikasi variabel penelitian

Kategori Variabel	Nama Variabel	Unit/Satuan
Variabel Independen (X)	Luas Panen	Hektar (ha)
	Curah Hujan	mm
	Suhu Udara	°C
	Kelembapan Udara	%
	Kecepatan Angin	m/s atau knot
Variabel Dependen (Y)	Total Produksi Padi	Ton

Alat dan Bahan

Alat yang digunakan pada penelitian ini terbagi menjadi dua bagian yakni perangkat keras dan lunak. Untuk perangkat kerasnya menggunakan laptop, *handphone*, dan perangkat pendukung dalam penelitian. Perangkat lunak yang digunakan meliputi bahasa pemrograman *Python* sebagai platform utama untuk analisis data, pengembangan model, dan visualisasi data. Proses tersebut didukung oleh berbagai pustaka *Python*, seperti *Pandas* untuk manipulasi data, *NumPy* untuk komputasi numerik, *Scikit-learn* untuk pengembangan model machine learning, serta *Matplotlib* untuk visualisasi data (Feng et al., 2022). Pengolahan data dilakukan menggunakan *Pandas* untuk pembersihan, transformasi, dan manipulasi data tabular, khususnya data iklim dan produksi padi (Choudhary et al., 2022). Implementasi penelitian dilakukan menggunakan bahasa pemrograman *Python* dengan memanfaatkan beberapa pustaka, yaitu *Pandas* untuk manipulasi data, *NumPy* untuk komputasi numerik, *Scikit-learn* untuk implementasi algoritma *machine learning Random Forest* dan *XGBoost*, serta *Matplotlib* untuk visualisasi data.

Bahan yang digunakan dalam penelitian ini berupa data yang digunakan adalah data sekunder periode 2018–2024 yang meliputi:

1. Data Agroklimat berupa curah hujan, suhu udara, kelembapan relatif, dan kecepatan angin periode 2018–2024 yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Data agroklimat dikompilasi untuk 27 kabupaten/kota di Provinsi Jawa Barat dengan menggunakan data dari stasiun BMKG yang mewakili masing-masing wilayah pengamatan (BMKG, 2025).
2. Data Produksi: Luas panen dan produktivitas padi per kabupaten/kota (BPS Jawa Barat, 2024)

Prosedur Penelitian

Tahapan penelitian diadaptasi dari kerangka *Knowledge Discovery in Databases* (KDD) (Hutahae et al., 2024). Model tersebut dipilih karena memiliki alur dalam membangun model prediksi berbasis data

Pengumpulan Data

Penelitian ini menggunakan data sekunder periode 2018–2024 yang terdiri atas luas panen, total produksi padi, curah hujan, suhu udara, kelembapan udara, dan kecepatan angin di Provinsi Jawa Barat. Data produksi dan luas panen diperoleh dari Badan Pusat Statistik (BPS), sedangkan data agroklimat diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Sebanyak 189 observasi digunakan sebagai dataset dalam pengembangan model prediksi. Variabel penelitian terlihat pada Tabel 1.

Pra-pemrosesan Data (Data Pre-processing)

Sebelum pemodelan, dilakukan pembersihan data untuk menjamin kualitas instrumen keputusan, meliputi:

1. Seleksi, dilakukan untuk menyaring data yang relevan, yang kemudian disesuaikan dengan kebutuhan analisis penelitian. Representasi data yang telah terpilih didokumentasikan pada Gambar 2.
2. Pembersihan data (*Cleaning*). Tahap pembersihan data dilakukan untuk memastikan kualitas data sebelum proses pemodelan. Penanganan *missing values* dilakukan dengan menghapus baris data yang memiliki nilai hilang karena jumlahnya relatif sedikit sehingga tidak memengaruhi representativitas data. Selanjutnya, deteksi *outlier* dilakukan menggunakan metode *Interquartile Range* (IQR) dengan menghitung batas bawah ($Q1 - 1,5 \times IQR$) dan batas atas ($Q3 + 1,5 \times IQR$). Data yang berada di luar rentang tersebut diidentifikasi sebagai *outlier* dan dihapus dari dataset agar distribusi data menjadi lebih representatif serta mengurangi pengaruh nilai ekstrem terhadap proses pelatihan model. (Raharusun & Hasibuan, 2024).
3. Transformasi data kategorikal pada fitur 'kabupaten/kota' menggunakan teknik *one-hot encoding*. Metode ini mengonversi data nominal menjadi variabel biner (*dummy*) untuk merepresentasikan kategori numerik.
4. Proses *feature scaling* dilakukan menggunakan metode *StandardScaler* untuk menstandarkan setiap fitur sehingga memiliki rata-rata (*mean*) sebesar 0 dan simpangan baku (*standard deviation*) sebesar 1. Proses ini bertujuan untuk menyetarakan skala antar fitur sehingga proses pelatihan model *Random Forest* dan *XGBoost* menjadi lebih stabil.
5. Tahap pembagian data (*data splitting*) dilakukan dengan memisahkan dataset menjadi dua bagian: 80% untuk proses pelatihan arsitektur model dan 20% untuk pengujian performa prediksi.

Implementasi Algoritme Ensemble Learning

Penelitian ini melakukan *benchmark* sistematis terhadap dua metode regresi berbasis *machine learning* untuk menangani data non linier dan multivariat:

1. Model *Random Forest* dibangun menggunakan sejumlah pohon keputusan (*decision trees*) yang dihasilkan melalui teknik *bootstrap sampling*. Pada proses pelatihan, model menggunakan 100 pohon keputusan ($n_estimators = 100$) dengan $random_state = 42$ untuk memastikan hasil pelatihan dapat direproduksi. Parameter lainnya menggunakan nilai bawaan (*default*) dari pustaka *Scikit-learn*. Penggunaan 100 pohon keputusan bertujuan meningkatkan stabilitas model melalui mekanisme *bagging* sehingga mampu mengurangi variansi prediksi dan risiko *overfitting* (Pratama & Astuti, 2025).
2. Model *XGBoost* dibangun menggunakan pendekatan *gradient boosting*, yaitu menyusun pohon keputusan secara bertahap untuk meminimalkan fungsi kerugian pada setiap iterasi. Pada penelitian ini, model menggunakan 100 iterasi ($n_estimators = 100$), $learning\ rate$ sebesar 0,1, kedalaman maksimum pohon (max_depth) sebesar 5, serta $random_state = 42$. Nilai *learning rate* digunakan untuk mengontrol besarnya pembaruan model pada setiap iterasi, sedangkan max_depth membatasi kompleksitas pohon keputusan agar model mampu mempelajari pola data tanpa meningkatkan risiko *overfitting*. Parameter lainnya menggunakan nilai bawaan (*default*) dari pustaka *XGBoost* (Sangaji & Sutabri, 2025).

Analisis dilakukan menggunakan bahasa pemrograman *Python* dengan dukungan pustaka *Pandas* untuk pengolahan data, *NumPy* untuk komputasi numerik, serta *Scikit-Learn* dan *XGBoost* untuk pembangunan model.

Evaluasi Model dan Keputusan

Validasi model dilakukan dengan membagi dataset menjadi data latih (*training set*) dan data uji (*test set*) menggunakan proporsi 80:20 setelah melalui tahap prapemrosesan data. Untuk mengevaluasi konsistensi kinerja model, penelitian ini menerapkan 5-fold cross-validation menggunakan fungsi `cross_val_score` pada pustaka *Scikit-learn*. Pada metode ini, dataset dibagi menjadi lima subset (*fold*) yang berukuran hampir sama. Pada setiap iterasi, empat *fold* digunakan sebagai data pelatihan dan satu *fold* sebagai data validasi. Proses ini diulang sebanyak lima kali sehingga setiap *fold* digunakan satu kali sebagai data validasi. Nilai validasi diperoleh dari rata-rata skor koefisien determinasi (R^2) pada kelima iterasi. Selanjutnya, model terbaik dievaluasi menggunakan data uji berdasarkan metrik R^2 , MAE, dan RMSE. Selanjutnya, efektivitas setiap model diukur berdasarkan tiga indikator kinerja statistik, yaitu:

1. Tingkat presisi model diestimasi melalui *Root Mean Square Error* (RMSE), yang merepresentasikan akar kuadrat dari rata-rata selisih kuadrat antara data observasi dan hasil estimasi. Nilai RMSE yang mendekati nol mengindikasikan tingkat kesalahan prediksi yang rendah dan akurasi yang lebih tinggi (Chai & Draxler, 2022). Rumus RMSE didefinisikan pada persamaan 1.

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

2. Untuk memperoleh gambaran mengenai besaran kesalahan rata-rata secara linear, digunakan *Mean Absolute Error* (MAE) yang mengkalkulasi rerata dari nilai absolut residu (Willmott et al., 2021). Berbeda dengan RMSE, MAE memberikan bobot yang sama terhadap seluruh selisih individu. Dapat dilihat pada persamaan 2.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

3. Kemampuan variabel prediktor (unsur iklim) dalam merepresentasikan variabilitas pada variabel respon (output padi) dinilai menggunakan Koefisien Determinasi (R^2). Metrik ini mengukur proporsi total varians data yang berhasil dijelaskan oleh model (Chicco et al., 2021). Rumusnya ditunjukkan pada persamaan 3.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} \quad (3)$$

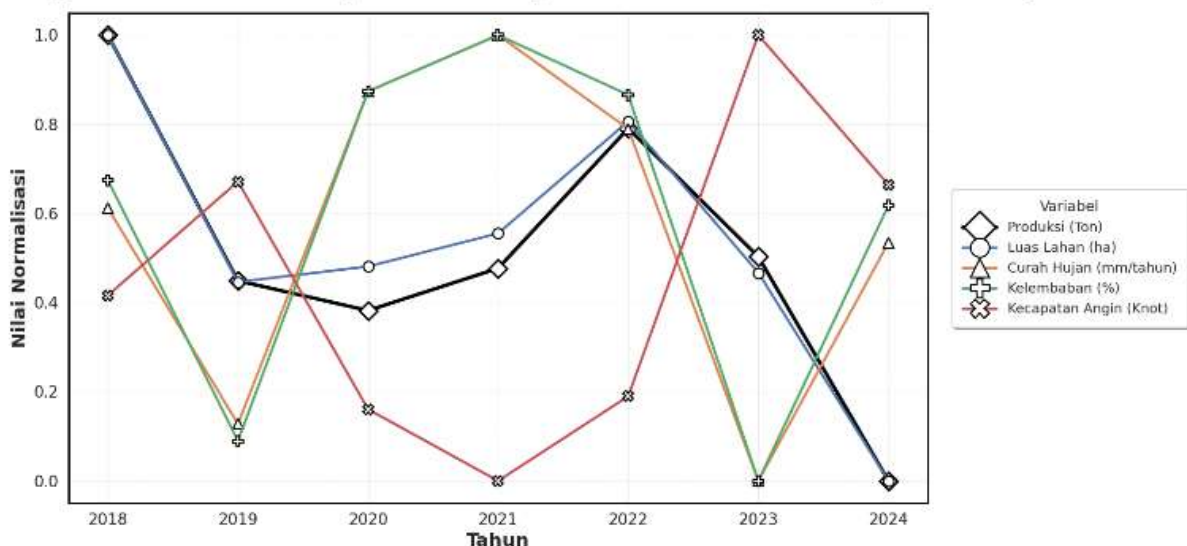
Keterangan:

- n : Ukuran sampel
 y_i : Data observasi produksi padi
 $\hat{y}_i$: Output model produksi padi
 $\bar{y}_i$: Rata-rata nilai aktual

Kabupaten/Kota	Tahun	Produksi (Ton)	Luas Lahan (ha)	Rata-rata Suhu (°C)	Kelembaban (%)	Curah Hujan (mm/tahun)	Kecapatan Angin (knot)	
0	Bandung	2018	314869.16	54427.68	23.0	81	2500	2.1
1	Bandung	2019	344214.39	56310.25	24.0	75	1950	2.0
2	Bandung	2020	277156.28	50150.32	23.0	82	2000	1.8
3	Bandung	2021	310715.20	50592.68	24.0	83	2950	1.9
4	Bandung	2022	289205.70	47537.79	24.0	82	2700	1.7
...	...	...	...	...	...	...	...	...
184	Tasikmalaya	2020	444358.24	72941.41	25.1	89	3450	2.7
185	Tasikmalaya	2021	445909.19	82934.85	25.0	96	3600	2.6
186	Tasikmalaya	2022	430961.73	81310.17	25.1	85	3350	2.8
187	Tasikmalaya	2023	391379.69	69685.72	26.3	77	2200	3.2
188	Tasikmalaya	2024	353415.07	66324.47	25.5	82	2950	3.0

Gambar 2. Data yang sudah diseleksi

Hubungan Produksi Padi dengan Variabel Agroklimat dan Luas Lahan (2018-2024)



Gambar 3. Korelasi tren produksi padi dengan *features*

HASIL DAN PEMBAHASAN

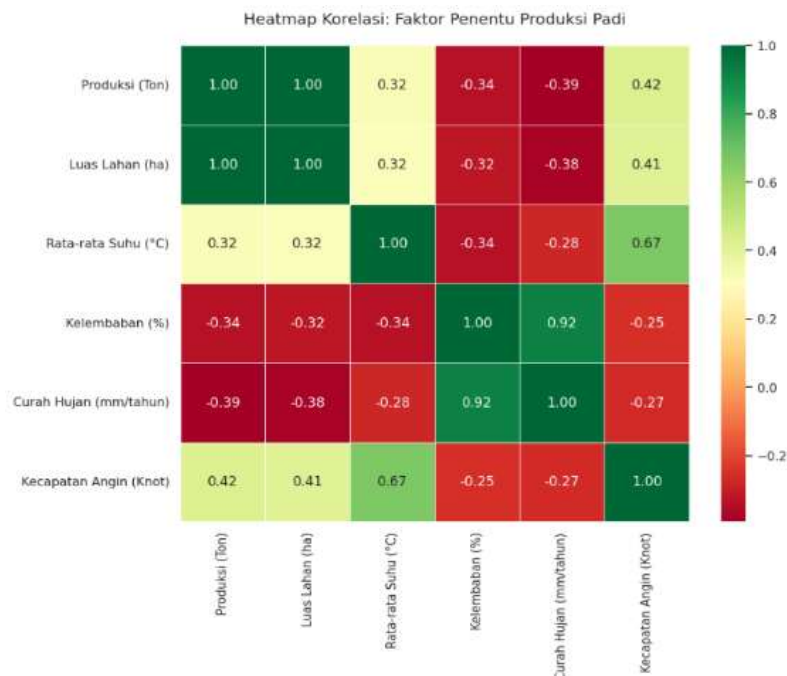
Karakteristik Data dan Analisis Deskriptif

Analisis dimulai dengan pengecekan terhadap dataset yang berisi 189 data dari berbagai Kabupaten/Kota di Provinsi Jawa Barat. Dataset ini menggabungkan produksi padi sebagai variabel utama dengan parameter iklim pertanian seperti suhu, kelembapan, curah hujan, dan kecepatan angin.

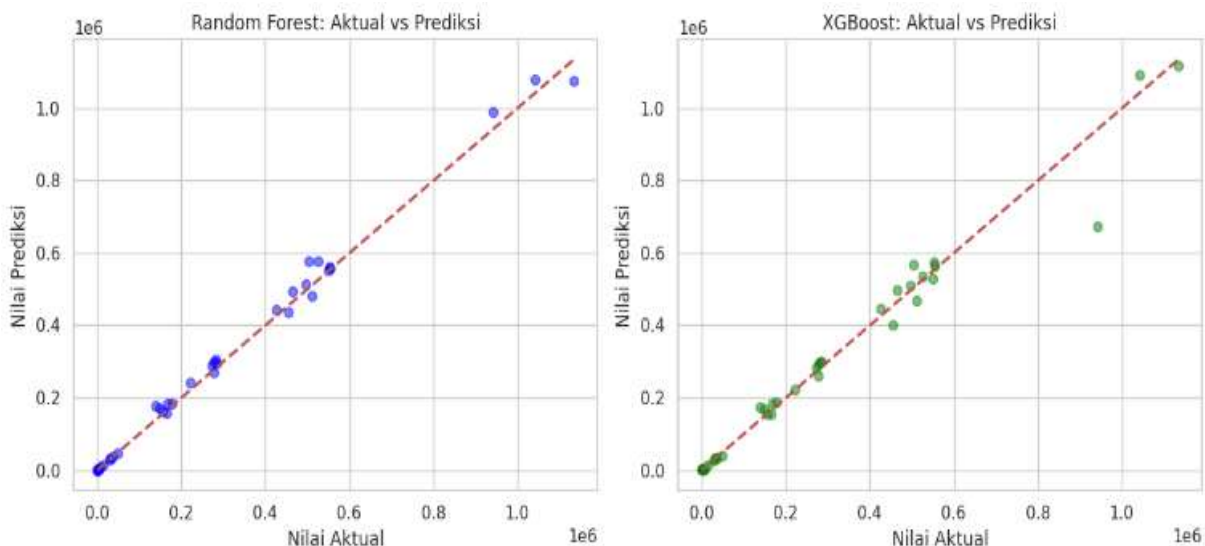
Berdasarkan ringkasan statistik pada Tabel 2, produksi padi di Provinsi Jawa Barat memiliki nilai rata-rata sebesar 338.959,3 ton dengan simpangan baku sebesar 361.751,8 ton, yang menunjukkan adanya variasi produksi yang cukup besar antar wilayah dan periode pengamatan. Luas lahan juga memiliki variasi yang tinggi dengan rata-rata 59.251,37 ha, sedangkan variabel agroklimat, seperti suhu udara, kelembapan, curah hujan, dan kecepatan angin, menunjukkan variasi yang relatif lebih kecil. Variasi karakteristik data tersebut mengindikasikan bahwa dataset memiliki keragaman yang cukup untuk digunakan

dalam pemodelan prediksi produksi padi. Selanjutnya, Gambar 3 memperlihatkan bahwa perubahan produksi padi dari tahun ke tahun diikuti oleh perubahan beberapa variabel agroklimat, sehingga mengindikasikan adanya hubungan antara kondisi agroklimat dan produksi padi. Temuan ini sejalan dengan Hussain et al. (2020) yang menyatakan bahwa kondisi iklim merupakan salah satu faktor yang memengaruhi produksi padi. Namun, penelitian ini menggunakan data tahunan sehingga belum dapat menjelaskan secara langsung pengaruh perubahan musiman terhadap variasi produksi padi.

Visualisasi *heatmap* korelasi (Gambar 4) mengonfirmasi bahwa meskipun luas lahan mendominasi secara linear, variabel agroklimat memiliki keterkaitan multivariat yang lebih kompleks. Hubungan non-linear ini menjadi justifikasi utama pemilihan algoritma *ensemble learning* seperti *Random Forest* dan *XGBoost*, yang mampu mengintegrasikan interaksi antar-variabel tersebut untuk menghasilkan estimasi produksi yang jauh lebih presisi dibandingkan metode statistik konvensional.



Gambar 4. Heatmap korelasi faktor penentu produksi padi



Gambar 5. Perbandingan model *random forest* dan *xgboost* terhadap nilai aktual dan prediksi produksi padi

Tabel 2. Statistik deskriptif variabel penelitian (n=189)

Variabel	Mean	Std. Dev	Min	Max
Produksi (Ton)	338.959,3	361.751,8	30,51	1.482.255,8
Luas Lahan (ha)	59.251,37	60.885,58	5,82	245.329,29
Rata-rata Suhu (°C)	26,17	1,79	22,0	29,1
Kelembapan (%)	79,18	4,58	68,0	88,0
Curah Hujan (mm/tahun)	2.516,93	773,17	1.050,0	4.750,0
Kecepatan Angin (Knot)	3,05	0,94	1,4	5,2

Sumber: Hasil pengolahan penulis (2026)

Tabel 3. Hasil komparasi efektivitas algoritma

Model	R ²	MAE	RMSE
<i>Random Forest</i>	0,992890	16.608,92	24.796,55
<i>XGBoost</i>	0,971003	22.410,47	50.075,24

Sumber: Hasil pengolahan penulis (2026)

Sebelum dilakukan proses pemodelan, analisis statistik deskriptif dilakukan untuk memberikan gambaran umum mengenai karakteristik data penelitian. Analisis ini mencakup nilai rata-rata (*mean*), simpangan baku (*standard deviation*), nilai minimum, dan nilai maksimum dari setiap variabel yang digunakan dalam penelitian. Informasi tersebut penting untuk memahami sebaran data serta variasi masing-masing variabel yang berpotensi memengaruhi hasil pemodelan. Hasil analisis statistik deskriptif disajikan pada Tabel 2.

Secara klimatologis, intensitas hujan di wilayah Jawa Barat tercatat sangat variatif, berkisar antara 1.050 hingga 4.750 mm per tahun. Sejalan dengan studi Elavarasan & Vincent (2021), fluktuasi curah hujan menjadi indikator kunci dalam membangun model estimasi hasil pertanian karena perannya dalam siklus hidrologi tanaman padi. Tingginya standar deviasi pada curah hujan dan luas lahan dalam penelitian ini mencerminkan adanya heterogenitas spasial di Jawa Barat, yang menurut Pratiwi *et al.* (2018), memerlukan pendekatan manajemen agribisnis yang spesifik lokasi guna mengantisipasi kegagalan panen.

Selain itu, suhu udara memiliki nilai rata-rata sebesar 26,17°C dan kelembapan sebesar 79,18%, yang menggambarkan karakteristik kondisi agroklimat pada wilayah penelitian. Kedua variabel tersebut digunakan sebagai variabel prediktor dalam model karena merupakan faktor iklim yang dapat memengaruhi pertumbuhan dan produktivitas padi (Yasin *et al.*, 2025). Pola data yang non-linear ini mengonfirmasi bahwa ketergantungan pada faktor alam di Jawa Barat sangat tinggi dan dari perspektif teknologi pertanian, karakteristik data ini menunjukkan pentingnya integrasi informasi agroklimat dalam pengembangan sistem prediksi produksi berbasis data.

Performa dan Perbandingan Model (Model Benchmarking)

Model *Random Forest* dan *XGBoost* dievaluasi menggunakan tiga metrik, yaitu koefisien determinasi (R^2), *Mean Absolute Error* (MAE), dan *Root Mean Square Error* (RMSE). Hasil perbandingan kinerja kedua algoritma disajikan pada Tabel 3.

Evaluasi model dilakukan untuk menilai kinerja prediksi algoritma yang digunakan dalam mengestimasi produksi padi di Provinsi Jawa Barat. Pada penelitian ini, dua algoritma *ensemble machine learning*, yaitu *Random Forest* (RF) dan *Extreme Gradient Boosting* (XGBoost), dibandingkan menggunakan tiga metrik evaluasi, yaitu koefisien determinasi (R^2), *Mean Absolute Error* (MAE), dan *Root Mean Square Error* (RMSE). Hasil perbandingan kinerja kedua algoritma disajikan pada Tabel 3. Berdasarkan hasil evaluasi tersebut, algoritma *Random Forest* menunjukkan performa yang lebih baik dibandingkan *XGBoost* pada dataset penelitian ini. Hal tersebut ditunjukkan oleh nilai koefisien determinasi (R^2) sebesar 0,992890, yang berarti 99,28% variasi produksi padi di Provinsi Jawa Barat dapat dijelaskan oleh variabel luas panen dan agroklimat yang digunakan dalam model. Selain itu, *Random Forest* menghasilkan nilai MAE sebesar 16.608,92 ton dan RMSE sebesar 24.796,55 ton, yang lebih rendah dibandingkan *XGBoost* dengan MAE sebesar 22.410,47 ton dan RMSE sebesar 50.075,24 ton. Nilai MAE dan RMSE yang lebih kecil menunjukkan bahwa hasil prediksi *Random Forest* lebih mendekati nilai aktual sehingga memiliki tingkat kesalahan prediksi yang lebih rendah. Temuan ini sejalan dengan penelitian Varma *et al.* (2024) yang melaporkan bahwa *Random Forest* mampu memberikan performa prediksi yang lebih baik pada beberapa kasus prediksi hasil panen karena lebih tangguh (*robust*) dalam menangani hubungan non linier dan variasi antar variabel. Namun demikian, performa kedua algoritma sangat dipengaruhi oleh karakteristik dataset, kualitas fitur, jumlah data, serta penyetelan *hyperparameter*. Oleh karena itu, hasil penelitian ini tidak dapat digeneralisasikan untuk seluruh kasus prediksi, karena pada dataset atau kondisi yang berbeda *XGBoost* dapat menghasilkan kinerja yang lebih baik dibandingkan *Random Forest* (Chen & Guestrin, 2016; Mahesh & Soundrapandian, 2024).

Selanjutnya, untuk melihat stabilitas model secara lebih mendalam, dilakukan analisis distribusi kesalahan menggunakan *boxplot* pada nilai *residual* (Gambar 6). Visualisasi *boxplot* menunjukkan bahwa model *Random Forest* sebarannya error terpusat secara simetris di sekitar garis nol dengan rentang interkuartil yang sempit. Hal ini mengindikasikan bahwa model memiliki varians yang stabil dan tidak menunjukkan adanya *error* ekstrem yang signifikan secara berulang. Dibandingkan dengan model *XGBoost* masih terdapat rentang yang cukup jauh dari garis nol.

Keunggulan *Random Forest* dalam penelitian ini selaras dengan temuan sebelumnya yang menyatakan bahwa algoritma berbasis *bagging* cenderung lebih tangguh (*robust*) terhadap variasi data pertanian regional yang memiliki korelasi multivariat yang kuat (Belgiu & Drăguț, 2016). Kemampuan *Random Forest* dalam merata-ratakan prediksi dari banyak pohon keputusan membantu meminimalkan risiko *overfitting* pada data agroklimat yang bersifat fluktuatif (Jeong *et al.*, 2016; Liakos *et al.*, 2018). Hal tersebut juga tercermin pada Gambar 6, di mana garis median (garis merah) *residual Random Forest* berada di sekitar nilai nol dengan sebaran *residual* yang lebih sempit dibandingkan *XGBoost*. Kondisi ini menunjukkan bahwa *Random Forest* menghasilkan prediksi yang lebih konsisten, memiliki bias prediksi yang lebih rendah, serta menghasilkan lebih sedikit *residual* ekstrem pada dataset penelitian ini. Temuan ini sejalan dengan Liakos *et al.* (2018) yang menyatakan bahwa penerapan *machine learning* pada bidang pertanian mampu meningkatkan akurasi prediksi dan efisiensi pengelolaan sumber daya. Namun demikian, hasil ini tidak dapat digeneralisasikan untuk semua kasus karena kinerja *Random Forest* maupun *XGBoost* sangat dipengaruhi oleh karakteristik data, kualitas fitur, jumlah

sampel, dan penyetelan *hyperparameter* (Elavarasan & Vincent, 2021; Chen & Guestrin, 2016).

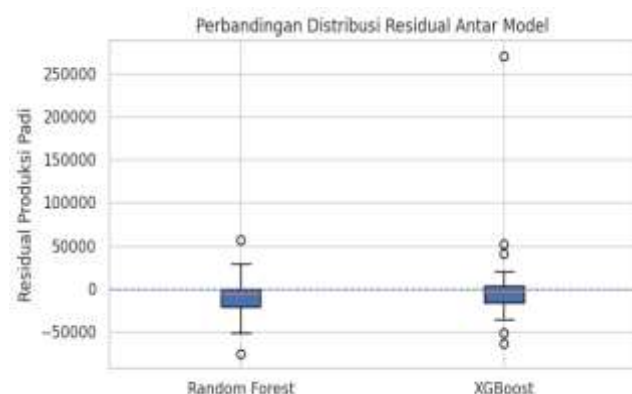
Analisis Kepentingan Fitur (*Feature Importance*)

Setelah menentukan bahwa *Random Forest* adalah model terbaik, dilakukan analisis kepentingan fitur untuk mengidentifikasi variabel agroklimat dan operasional yang paling dominan dalam menentukan hasil produksi padi di Jawa Barat. Analisis ini sangat krusial dalam perspektif agribisnis untuk menentukan prioritas mitigasi risiko. Hasil analisis fitur disajikan pada Gambar 7.

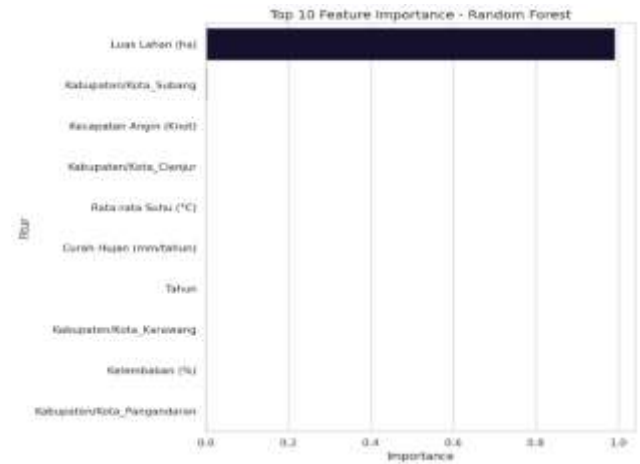
Berdasarkan hasil analisis *feature importance*, variabel luas lahan memberikan kontribusi terbesar terhadap prediksi produksi padi. Temuan ini konsisten dengan teori produksi pertanian yang menyatakan bahwa luas lahan merupakan faktor produksi utama yang menentukan kapasitas output tanaman. Selain itu, variabel agroklimat seperti curah hujan, suhu, kelembapan, dan kecepatan angin juga memiliki kontribusi dalam model, meskipun nilainya relatif lebih kecil dibandingkan luas lahan. Hasil ini menunjukkan bahwa prediksi produksi padi tidak hanya dipengaruhi oleh faktor luas lahan, tetapi juga mempertimbangkan kondisi agroklimat sebagai variabel pendukung, sebagaimana dijelaskan oleh Juliyanti dan Usman (2018) bahwa kondisi lingkungan berperan dalam proses fisiologis tanaman padi.

Hasil analisis *feature importance* yang disajikan pada Gambar 7 dan Tabel 4 menunjukkan bahwa variabel luas lahan merupakan prediktor utama dalam model dengan nilai *feature importance* sebesar 0,9914 pada *Random Forest* dan 0,9101 pada *XGBoost*. Meskipun demikian, variabel agroklimat, seperti kecepatan angin, suhu udara, curah hujan, dan kelembapan, tetap memberikan kontribusi terhadap proses prediksi, meskipun nilainya relatif lebih kecil dibandingkan luas lahan. Temuan ini menunjukkan bahwa estimasi produksi padi tidak hanya dipengaruhi oleh ketersediaan lahan, tetapi juga mempertimbangkan informasi agroklimat sebagai variabel pendukung dalam model prediksi.

Hasil tersebut sejalan dengan Elavarasan dan Vincent (2021), yang menyatakan bahwa selain faktor luas lahan, kondisi termal dan hidrologi selama masa pertanian turut memengaruhi hasil produksi tanaman. Dalam konteks pengelolaan pertanian di Provinsi Jawa Barat, integrasi variabel agroklimat ke dalam model prediksi memberikan informasi yang dapat dimanfaatkan untuk mendukung pengambilan keputusan, seperti perencanaan waktu tanam, pengelolaan irigasi, serta mitigasi risiko akibat anomali iklim pada daerah sentra produksi padi. Kebaruan penelitian ini terletak pada pengembangan model prediksi produksi padi berbasis integrasi data agroklimat multivariat pada skala regional menggunakan pendekatan *ensemble machine learning*.



Gambar 6. Perbandingan distribusi residual antar model



Gambar 7. Tingkat Kepentingan Variabel (*Feature Importance*) dalam Prediksi Produksi Padi

Tabel 4. Nilai *Feature Importance* Variabel pada Model *Random Forest* dan *XGBoost*

No	Variabel	<i>Random Forest</i>	<i>XGBoost</i>
1	Luas Lahan (ha)	0,9914	0,9101
2	Kabupaten/Kota Subang	0,0024	0,0000
3	Kecepatan Angin (Knot)	0,0017	0,0066
4	Kabupaten/Kota Cianjur	0,0009	0,0634
5	Rata-rata Suhu (°C)	0,0008	0,0012
6	Curah Hujan (mm/tahun)	0,0006	0,0004
7	Tahun	0,0006	0,0003
8	Kabupaten/Kota Karawang	0,0005	0,0000
9	Kelembapan (%)	0,0004	0,0017
10	Kabupaten/Kota Pangandaran	0,0002	0,0000

Berbeda dengan penelitian sebelumnya yang umumnya berfokus pada analisis statistik atau prediksi berbasis variabel produksi konvensional, penelitian ini mengintegrasikan variabel produksi dan agroklimat ke dalam satu kerangka prediktif menggunakan algoritma *Random Forest* dan *XGBoost*, sehingga berpotensi diimplementasikan sebagai teknologi pendukung pengelolaan produksi pertanian berbasis data.

Meskipun hasil analisis *feature importance* menunjukkan bahwa variabel luas lahan merupakan prediktor paling dominan dengan nilai *feature importance* sebesar 0,9914 pada *Random Forest* dan 0,9101 pada *XGBoost*, hal tersebut tidak berarti bahwa variabel agroklimat tidak berperan dalam proses prediksi. Dominannya luas lahan menunjukkan bahwa variasi produksi padi pada dataset penelitian ini lebih banyak dipengaruhi oleh perbedaan luas lahan antar wilayah. Sementara itu, variabel agroklimat, seperti curah hujan, suhu udara, kelembapan, dan kecepatan angin, memiliki kontribusi relatif yang lebih kecil, tetapi tetap dipertahankan dalam model sebagai variabel pendukung yang melengkapi informasi produksi. Integrasi variabel agroklimat memungkinkan model mempertimbangkan kondisi lingkungan dalam proses prediksi sehingga menghasilkan model yang lebih komprehensif dibandingkan apabila hanya menggunakan variabel produksi konvensional.

Dengan demikian, kebaruan penelitian ini tidak terletak pada dominannya pengaruh variabel agroklimat terhadap produksi padi, melainkan pada pengembangan model

prediksi yang mengintegrasikan data produksi dan data agroklimat dalam satu kerangka *ensemble machine learning* untuk mendukung prediksi produksi padi berbasis data pada skala regional. Namun demikian, hasil penelitian ini juga menunjukkan bahwa kontribusi variabel agroklimat masih relatif kecil dibandingkan variabel luas lahan. Kondisi tersebut mengindikasikan bahwa informasi agroklimat yang digunakan pada penelitian ini belum sepenuhnya mampu menjelaskan variasi produksi padi, sehingga diperlukan pengembangan model dengan menambahkan variabel lingkungan dan budidaya yang lebih representatif.

Oleh karena itu, penelitian ini memiliki beberapa keterbatasan. Pertama, data yang digunakan merupakan data sekunder dengan cakupan wilayah terbatas pada Provinsi Jawa Barat sehingga generalisasi model pada wilayah lain masih memerlukan validasi lebih lanjut. Kedua, variabel penelitian belum mencakup faktor manajemen budidaya, seperti penggunaan pupuk, varietas tanaman, sistem irigasi, kondisi tanah, serta praktik agronomi lainnya yang berpotensi memengaruhi produktivitas padi. Oleh karena itu, penelitian selanjutnya disarankan mengintegrasikan variabel teknis budidaya, data penginderaan jauh (*remote sensing*), maupun indikator agroklimat yang lebih rinci serta melakukan validasi lapangan untuk meningkatkan akurasi model dan mendukung implementasi sistem prediksi dalam skala operasional.

KESIMPULAN

Penelitian ini mengembangkan model estimasi produksi padi berbasis integrasi variabel agroklimat menggunakan pendekatan *ensemble machine learning* di Provinsi Jawa Barat. Berbeda dengan penelitian sebelumnya yang menggunakan variabel terbatas, penelitian ini mengintegrasikan data agroklimat multivariat serta membandingkan algoritma *Random Forest* dan *XGBoost* untuk meningkatkan akurasi prediksi. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* memberikan performa terbaik dengan nilai koefisien determinasi (R^2) sebesar 0,9928, *Mean Absolute Error* (MAE) sebesar 16.608,93 ton, dan *Root Mean Square Error* (RMSE) sebesar 24.796,56 ton. Nilai tersebut menunjukkan bahwa model *Random Forest* mampu menjelaskan 99,28% variasi produksi padi serta menghasilkan tingkat kesalahan prediksi yang lebih rendah dibandingkan *XGBoost*, sehingga prediksi yang dihasilkan lebih mendekati nilai aktual. Analisis kepentingan fitur menunjukkan bahwa variabel luas lahan merupakan faktor yang paling dominan dalam memengaruhi prediksi produksi padi pada model *Random Forest*. Nilai *feature importance* luas lahan jauh lebih tinggi dibandingkan variabel lainnya, seperti curah hujan, suhu udara, kelembapan, kecepatan angin, maupun variabel kabupaten/kota. Hasil ini mengindikasikan bahwa luas lahan memberikan kontribusi terbesar terhadap kemampuan model dalam memprediksi produksi padi pada dataset penelitian ini, sedangkan variabel agroklimat berperan sebagai faktor pendukung dengan tingkat kontribusi yang relatif lebih kecil. Secara aplikatif, model ini berpotensi dimanfaatkan sebagai komponen pendukung keputusan dalam perencanaan produksi, pengelolaan rantai pasok, serta pengembangan *precision agriculture* berbasis data.

Penelitian ini memiliki beberapa keterbatasan. Pertama, data yang digunakan merupakan data sekunder dengan cakupan wilayah terbatas pada Provinsi Jawa Barat sehingga generalisasi model ke wilayah lain masih memerlukan validasi lebih lanjut. Kedua, variabel penelitian belum mencakup faktor manajemen budidaya, seperti penggunaan pupuk, varietas tanaman, sistem irigasi, jenis

tanah, dan praktik agronomi yang berpotensi memengaruhi produktivitas padi. Ketiga, penelitian ini hanya membandingkan dua algoritma *ensemble machine learning*, yaitu *Random Forest* dan *XGBoost*, sehingga belum mengevaluasi performanya terhadap algoritma lain maupun pendekatan pemodelan spasial yang mampu mempertimbangkan heterogenitas dan ketergantungan spasial antar wilayah. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas cakupan wilayah penelitian hingga tingkat nasional, memperpanjang periode pengamatan, serta mengintegrasikan data penginderaan jauh, karakteristik tanah, dan praktik budidaya agar model mampu merepresentasikan kondisi produksi padi secara lebih komprehensif. Selain itu, penelitian mendatang dapat membandingkan algoritma lain, seperti *LightGBM*, *CatBoost*, *Support Vector Regression*, maupun model *deep learning* berbasis *Long Short-Term Memory* (LSTM), serta mengembangkan pendekatan pemodelan spasial, seperti *Geographically Weighted Regression* (GWR), *Geographically Weighted Random Forest* (GWRF), atau metode spasial lainnya. Pengembangan lanjutan juga dapat diarahkan pada implementasi model ke dalam sistem pendukung keputusan atau aplikasi berbasis web yang mampu menghasilkan prediksi produksi padi secara otomatis berdasarkan data agroklimat terkini, sehingga dapat dimanfaatkan oleh pemerintah, penyuluh pertanian, dan pemangku kepentingan dalam mendukung perencanaan produksi serta pengambilan kebijakan di sektor pertanian.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada BPS Provinsi Jawa Barat dan BMKG atas penyediaan data penelitian, serta kepada Fakultas Pertanian Universitas Singaperbangsa Karawang atas dukungan akademik selama penelitian ini berlangsung.

DAFTAR PUSTAKA

- Ariyani, N., & Arifin, A. Z. (2021). Prediksi tingkat pengangguran di Kabupaten Tuban tahun 2020 menggunakan metode regresi linear sederhana. *Mathvision*, 3(1), 6–13. <https://doi.org/10.55719/mv.v3i1.245>
- Badan Meteorologi, Klimatologi, dan Geofisika. (2025). *Data Curah Hujan Harian Wilayah Jawa Barat*. <https://data.bmkg.go.id/>
- Badan Pusat Statistik Provinsi Jawa Barat. (2024). *Produksi Tanaman Padi Menurut Kabupaten/Kota di Provinsi Jawa Barat*. Diakses pada 21 Desember 2025 dari <https://jabar.bps.go.id/indicator/>
- Basahona, A. A., Ishak, R., & Husna, A. (2019). Penerapan metode linier regresi untuk prediksi produksi sayur-sayuran. *Jurnal Nasional Cosphi*, 3(2), 54–57.
- Belgiu, M., & Drăguț, L. (2016). *Random Forest* in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE). *Geoscientific model development discussions*, 7(1), 1525–1534
- Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge

- Discovery and Data Mining (KDD '16), 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chicco, D., Warrens, M. J., & Jurman, G. (2021). *The coefficient of determination R-squared is more informative than SMAPE, MAE, and RMSE in regression analysis evaluation*. *PeerJ Computer Science*, 7, e623. <https://doi.org/10.7717/peerj-cs.623>
- Choudhary, K., Shi, W., Dong, Y., & Paringer, R. (2022). *Random Forest for rice yield mapping and prediction using Sentinel-2 data with Google Earth Engine*. *Advances in Space Research*, 70(8), 2443–2457. <https://doi.org/10.1016/j.asr.2022.07.016>
- Elavarasan, D., & Vincent, P. D. R. (2021). A reinforced *Random Forest* model for enhanced crop yield prediction by integrating agrarian parameters. *Journal of Ambient Intelligence and Humanized Computing*, 12(11), 10009–10022. <https://doi.org/10.1007/s12652-020-02753-y>
- FAO. (2021). *The state of food and agriculture 2021: Making agrifood systems more resilient to shocks and stresses*. Food and Agriculture Organization of the United Nations. <https://www.fao.org/3/cb4476en/cb4476en.pdf>
- Feng, P., Wang, B., Liu, D. L., Waters, C., Xiao, D., Shi, L., & Yu, Q. (2022). Evaluation of *Random Forests* (RF) for regional and local-scale wheat yield prediction in southeast Australia. *Sensors*, 22(3), 717. <https://doi.org/10.3390/s22030717>
- Gunawan, G. P., Ruminta, R., & Wicaksono, F. Y. (2024). Identifikasi Perubahan Iklim dan Korelasinya terhadap Produksi Tanaman Padi (*Oryza sativa* L.) di Kabupaten Karawang. *Paspalum: Jurnal Ilmiah Pertanian*, 12(2), 428-433.
- Herwanto, H. W., Widiyaningtyas, T., & Indriana, P. (2019). Penerapan algoritme linear regression untuk prediksi hasil panen tanaman padi. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, 8(4), 364–370. <https://doi.org/10.22146/jnteti.v8i4.537>
- Hussain, S., Huang, J., Huang, J., Ahmad, S., Nanda, S., Anwar, S., ... & Zhang, J. (2020). Rice production under climate change: adaptations and mitigating strategies. In *Environment, climate, plant and vegetation growth* (pp. 659-686). Cham: Springer International Publishing.
- Hutahaean, J., Yusup, D., & Purwantoro, P. (2024). Perbandingan Metode Linear Regression, *Random Forest* & K-Nearest Neighbor Untuk Prediksi Produksi Hasil Panen Padi Di Provinsi Jawa Barat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3895-3900.
- Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K. M., & Gerber, J. S. (2016). *Random Forests* for global and regional crop yield predictions. *PLoS ONE*, 11(6), e0156571. <https://doi.org/10.1371/journal.pone.0156571>
- Juliyanti, J., & Usman, U. (2018). Pengaruh luas lahan, pupuk dan jumlah tenaga kerja terhadap produksi padi gampong matang baloi. *Jurnal Ekonomi Pertanian Unimal*, 1(1), 31-39.
- Kharisma S, M. I., Hadiana, A. I., & Ramadhan, E. (2025). Model Prediksi Produksi Padi Berdasarkan Curah Hujan dan Suhu Menggunakan Regresi Linier Berganda. *Jurnal Algoritma*, 22(2). <https://doi.org/10.33364/algoritma/v.22-2.2793>
- Li, X., Wang, Y., & Chen, J. (2023). Forecast-driven inventory management for agricultural supply chains under climate uncertainty. *Computers and Electronics in Agriculture*, 204, 107531. <https://doi.org/10.1016/j.compag.2023.107531>
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674. <https://doi.org/10.3390/s18082674>
- Mahesh, P., & Soundrapandiyam, R. (2024). *Yield prediction for crops by gradient-based algorithms*. *PLOS ONE*, 19(8), e0291928. <https://doi.org/10.1371/journal.pone.0291928>
- Pratama, Z. B. P., & Astuti, Y. P. (2025). Comparison of Machine Learning Methods (Linear Regression, *Random Forest*, and XGBoost) for Predicting Poverty in Central Java in 2024. *Sistemasi: Jurnal Sistem Informasi*, 14(5), 2492-2499.
- Pratiwi, E. E., Widodo, A. W., & Mahmudy, W. F. (2018). Penerapan Algoritme Genetika Pada Kasus Optimasi Penentuan Bibit dan Pemerataan Subsidi pupuk (Studi Kasus: Desa Pandansari, Kabupaten Kediri). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(5), 1803-1812.
- Raharusun, O. S. H., & Hasibuan, A. (2024). Sistem Prediksi Produksi Beras Menggunakan Multiple Linear Regression untuk Optimalisasi Ketahanan Pangan di Kabupaten Minahasa. *JOINTER: Journal of Informatics Engineering*, 5(02), 36-40.
- Saadah, S., & Salsabila, H. (2021). Prediksi harga bitcoin menggunakan metode *Random Forest*: (Studi kasus: data acak pada masa pandemic covid-19). *Jurnal Komputer Terapan*, 7(1), 24–32. <https://doi.org/10.35143/jkt.v7i1.4632>
- Sangaji, D., & Sutabri, T. (2025). Analisis XGBoost dan *Random Forest* untuk Prediksi Curah Hujan dalam Mendukung Mitigasi Karhutla. *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*, 5(1), 13-18. <https://doi.org/10.55382/jurnalpustakaai.v5i1.905>
- Sulardi, P., Hendro, T., & Umbara, F. R. (2017). Prediksi Kebutuhan Obat Menggunakan Regresi Linier. In *Seminar Nasional Teknologi dan Informatika 2017*. Muria Kudus University.
- Varma, M., Kumar, P. A., Yadav, K. K., Verma, S., Kumar, P., & Singh, A. K. (2024). A comparison of *Random Forest* based models for agricultural crop yield prediction. *International Journal of Research in Agronomy*, 7(8S), 817–820. <https://doi.org/10.33545/2618060X.2024.v7.i8Sk.1386>
- Willmott, C. J., Matsuura, K., & Robeson, S. M. (2021). Ambiguities in the application and interpretation of root-mean-square error (RMSE) and mean absolute error (MAE). *Annals of the American Association of Geographers*, 111(1), 212-219.
- Yasin, F., Firmansyah, M. R. U., Aldo, D., & Amrustian, M. A. (2025). Multivariate Forecasting of Paddy Production: A Comparative Study of Machine Learning Models. *Jurnal Teknik Informatika (Jutif)*, 6(3), 1431-1442.
- Zhang, Y., Zhao, L., & Liu, H. (2022). Crop yield prediction and its implications for supply chain planning in agriculture. *Agricultural Systems*, 198, 103377. <https://doi.org/10.1016/j.agry.2022.103377>

Halaman ini sengaja dikosongkan