

Prediksi Hasil Panen untuk Pertanian Menggunakan Model Regresi Machine Learning

Arig Kusuma Jati¹, Bayu Rizky Utomo², Naufal Hanan Jati Asmara³,
Rajnaparamitha Kusumastusi⁴

¹²³Prodi S1 Informatika, ⁴Prodi S1 Teknologi Informasi, STMIK Amikom Surakarta
¹²³⁴Sukoharjo Indonesia

Email: ¹arig.10425@mhs.amikomsolo.ac.id, ²bayu.10423@mhs.amikomsolo.ac.id,
³naufal.10413@mhs.amikomsolo.ac.id, ⁴rajna@dosen.amikomsolo.ac.id

Abstract

Agricultural productivity is influenced by climatic variables, agronomic conditions, and the use of cultivation inputs, making accurate crop yield prediction a critical challenge in the implementation of precision agriculture. Although previous studies have applied various machine learning models to estimate crop yields, most research has not compared the performance of multiple regression algorithms simultaneously on a multi-country dataset that includes diverse commodities and highly variable environmental conditions. This gap highlights the need for a comprehensive analysis to identify the most stable and accurate model for yield prediction. This study aims to evaluate and compare the performance of three machine learning regression models—Decision Tree Regressor, Random Forest Regressor, and Gradient Boosting Regressor—in predicting crop yields based on area, crop type, rainfall, temperature, and pesticide usage. The methodology involves data preprocessing (removal of irrelevant columns, categorical encoding, and numerical feature standardization), outlier detection and removal using the Z-Score method, an 80:20 train–test split, and hyperparameter optimization using GridSearchCV with k-fold cross-validation. The dataset used consists of historical global agricultural yield records encompassing diverse climates, commodities, and regions. The results show that the Random Forest Regressor achieves the best performance, with an MAE of 3053.90 and an R^2 of 0.9860 after applying One-Hot Encoding, outlier removal, and hyperparameter tuning. These findings indicate that tree-based ensemble models provide higher prediction stability compared to single-tree or boosting methods. This study contributes by identifying the most effective model for large-scale yield prediction and demonstrating the potential of machine learning in supporting decision-making processes within precision agriculture.

Keywords: crop yield prediction, machine learning, Random Forest, regression, precision agriculture

Abstraksi

Produktivitas pertanian dipengaruhi oleh variabel iklim, kondisi agronomi, serta penggunaan input budidaya, sehingga akurasi prediksi hasil panen menjadi tantangan penting dalam penerapan pertanian presisi. Meskipun berbagai studi sebelumnya telah menerapkan model machine learning untuk estimasi hasil panen, sebagian besar penelitian belum membandingkan performa beberapa algoritma regresi sekaligus pada

dataset multinegara dengan keragaman komoditas dan kondisi lingkungan yang tinggi. Kesenjangan ini menunjukkan perlunya analisis komprehensif untuk mengidentifikasi model yang paling stabil dan akurat dalam memprediksi hasil panen. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja tiga model regresi machine learning—Decision Tree Regressor, Random Forest Regressor, dan Gradient Boosting Regressor—dalam memprediksi hasil panen berdasarkan variabel area, jenis tanaman, curah hujan, temperatur, dan penggunaan pestisida. Metode penelitian meliputi preprocessing data (penghapusan kolom tidak relevan, encoding kategorikal, standarisasi fitur numerik), deteksi dan penghapusan outlier menggunakan Z-Score, pembagian data latih-uji 80:20, serta optimasi hyperparameter menggunakan GridSearchCV dengan k-fold cross-validation. Dataset yang digunakan merupakan data historis yield agrikultur global dengan berbagai variasi iklim, komoditas, dan wilayah. Hasil penelitian menunjukkan bahwa Random Forest Regressor menghasilkan performa terbaik dengan nilai MAE sebesar 3053,90 dan R^2 sebesar 0,9860 setelah penerapan One-Hot Encoding, penghapusan outlier, dan tuning hyperparameter. Temuan ini menunjukkan bahwa model ensemble berbasis pohon memberikan stabilitas prediksi yang lebih tinggi dibandingkan model tunggal maupun boosting. Penelitian ini memberikan kontribusi berupa identifikasi model paling efektif untuk prediksi hasil panen pada skala wilayah luas serta menunjukkan potensi penerapan machine learning dalam mendukung pengambilan keputusan pada pertanian presisi.

Kata kunci: prediksi hasil panen, machine learning, Random Forest, regresi, pertanian presisi

1. PENDAHULUAN

Sektor pertanian memiliki peran strategis dalam menjaga ketahanan pangan dan mendukung perekonomian nasional, terutama di negara agraris seperti Indonesia. Produktivitas pertanian dipengaruhi oleh berbagai faktor lingkungan dan agronomi, seperti kondisi iklim, jenis tanaman, kualitas tanah, penggunaan input pertanian, serta manajemen budidaya [1], [2]. Variasi kompleks dari faktor-faktor tersebut menyebabkan hasil panen sering kali sulit diprediksi secara akurat. Ketidakpastian ini berdampak pada proses perencanaan distribusi, pengelolaan rantai pasok, serta optimalisasi sumber daya pertanian [3].

Perubahan iklim global turut memperbesar kompleksitas permasalahan melalui pergeseran pola curah hujan dan peningkatan suhu rata-rata yang mempengaruhi stabilitas produksi tanaman pangan [4]. Di tengah kondisi tersebut, pendekatan prediksi tradisional seperti perhitungan rata-rata historis atau model statistik linear memiliki keterbatasan dalam menangkap hubungan non-linear antar variabel agronomi dan iklim—sebuah fenomena yang banyak ditemukan pada sistem pertanian modern [5]. Oleh karena itu, metode analitik yang mampu memodelkan pola kompleks secara lebih adaptif menjadi semakin dibutuhkan.

Perkembangan *machine learning* membuka peluang signifikan dalam peningkatan akurasi prediksi hasil panen. Algoritma *machine learning* mampu mengolah data dalam jumlah besar serta mengidentifikasi pola yang tidak dapat ditangkap oleh

metode tradisional [6]. Berbagai studi sebelumnya menunjukkan bahwa model berbasis pohon keputusan dan metode ensemble seperti *Random Forest* dan *Gradient Boosting* mampu memberikan performa prediksi yang unggul pada berbagai komoditas pertanian [1], [7], [8]. Temuan tersebut menunjukkan bahwa pendekatan berbasis *machine learning* berpotensi menjadi alat pendukung keputusan yang efektif dalam perencanaan produksi dan mitigasi risiko agrikultur.

Berdasarkan konteks tersebut, penelitian ini bertujuan untuk mengembangkan model prediksi hasil panen menggunakan tiga algoritma regresi—*Decision Tree Regressor*, *Random Forest Regressor*, dan *Gradient Boosting Regressor*. Pemilihan ketiga model ini didasarkan pada kemampuannya dalam menangani data heterogen, sifatnya yang robust terhadap hubungan non-linear, serta efektivitasnya dalam berbagai penelitian terdahulu mengenai prediksi hasil pertanian. Penelitian ini juga berfokus pada perbandingan performa antar model untuk menentukan algoritma dengan tingkat akurasi dan stabilitas terbaik dalam memprediksi hasil panen.

2. TINJAUAN PUSTAKA

Penelitian mengenai prediksi hasil panen berbasis *machine learning* telah mengalami perkembangan yang signifikan seiring meningkatnya kebutuhan akan sistem pendukung keputusan di sektor pertanian. Sebagian besar penelitian terdahulu menekankan bahwa variabilitas faktor lingkungan, iklim, dan karakteristik agronomi membuat hasil panen bersifat non-linear dan sulit diprediksi secara akurat dengan pendekatan tradisional. Model statistik konvensional seperti regresi linear sering kali menunjukkan keterbatasan karena hanya mampu memodelkan hubungan linier dan tidak efektif ketika data memiliki interaksi multivariat yang kompleks [1], [3]. Kondisi ini mendorong penggunaan algoritma *machine learning* yang lebih adaptif.

Salah satu pendekatan yang paling banyak digunakan adalah metode berbasis pohon keputusan. Model *Decision Tree Regressor* sering menjadi titik awal dalam berbagai penelitian karena mampu menangani fitur kategorikal maupun numerik serta memberikan interpretasi yang jelas. Namun, penelitian sebelumnya menunjukkan bahwa *decision tree* cenderung rentan terhadap overfitting, terutama ketika data pelatihan memiliki keragaman yang tinggi atau distribusi yang tidak merata [5]. Untuk mengatasi keterbatasan tersebut, penelitian agrikultur modern mulai beralih ke metode ensemble.

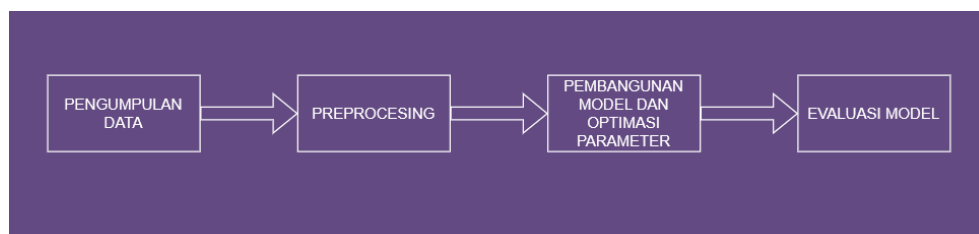
Metode ensemble seperti *Random Forest Regressor* banyak digunakan untuk prediksi hasil panen karena menggabungkan banyak pohon keputusan sehingga mampu meningkatkan stabilitas model dan mengurangi risiko overfitting. Sejumlah penelitian menunjukkan bahwa *Random Forest* menghasilkan akurasi lebih tinggi dibandingkan model regresi tradisional maupun *decision tree* tunggal pada berbagai komoditas pertanian [7], [9]. Keunggulan utama model ini terletak pada kemampuannya menangani fitur heterogen dan interaksi non-linear secara lebih efektif.

Sementara itu, *Gradient Boosting Regressor* dikembangkan dengan pendekatan boosting, yaitu memperbaiki kesalahan prediksi secara bertahap dari model sebelumnya. Sejumlah studi terkait prediksi tanaman biofarmaka dan hortikultura menunjukkan bahwa metode boosting dapat memberikan performa yang sangat baik ketika parameter dioptimasi dengan benar [10]. Meskipun demikian, model ini sensitif terhadap konfigurasi hyperparameter dan dapat mengalami underfitting atau overfitting jika tidak ditangani dengan tepat.

Berdasarkan telaah literatur tersebut, pemilihan *Decision Tree Regressor*, *Random Forest Regressor*, dan *Gradient Boosting Regressor* dalam penelitian ini memiliki dasar yang jelas. *Decision Tree* digunakan sebagai model dasar untuk memahami struktur hubungan antar variabel secara interpretatif. *Random Forest* dipilih karena terbukti memiliki stabilitas tinggi dan performa yang konsisten pada data pertanian yang bersifat heterogen. Sedangkan *Gradient Boosting* dipilih karena kemampuannya mempelajari pola kompleks secara bertahap sehingga berpotensi menghasilkan akurasi lebih tinggi dalam kondisi parameter yang optimal. Dengan demikian, tinjauan pustaka ini secara logis mengarah pada alasan pemilihan ketiga model sebagai fokus utama penelitian.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen yang dirancang untuk mengembangkan model prediksi hasil panen berbasis *machine learning*. Jenis penelitian ini termasuk dalam kategori penelitian terapan (*applied research*) karena fokus utamanya adalah implementasi algoritma untuk memecahkan masalah nyata di sektor pertanian. Desain penelitian bersifat eksperimental komparatif dengan tujuan membandingkan performa beberapa algoritma regresi sehingga dapat diidentifikasi model dengan akurasi prediksi tertinggi dan stabilitas hasil yang baik. Tahapan penelitian disusun secara sistematis mengikuti alur pada Gambar 1, yaitu mulai dari pengumpulan data, *preprocessing*, pembangunan model, optimasi parameter, hingga evaluasi performa. Berikut ini adalah alur penelitiannya yang ditampilkan pada Gambar 1.



Gambar 1. Alur Penelitian

3.1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini adalah FAOSTAT Crop Production Dataset yang dirilis oleh Food and Agriculture Organization (FAO) pada tahun 2023. Dataset ini berisi informasi produksi tanaman dari berbagai negara dengan variabel yang mencakup *Area*, *Item*, *Year*, *avg_temp*, *avg_rainfall*, dan *Yield*.

Setelah dilakukan seleksi fitur, dataset yang digunakan memiliki karakteristik sebagai berikut: Jumlah negara: 174 negara, Jumlah komoditas pertanian: 52 komoditas, Rentang tahun: 1990–2020, Jumlah total observasi: 10.231 baris data. Distribusi data bersifat heterogen karena setiap negara memiliki tingkat produksi, iklim, dan komoditas yang berbeda. Variabilitas yang tinggi ini membuat analisis *machine learning* menjadi relevan karena mampu menangani hubungan non-linear serta interaksi multivariat yang kompleks. Tahap eksplorasi awal dilakukan melalui histogram, *density plot*, dan *heatmap* korelasi. Visualisasi ini menunjukkan bahwa variabel *Yield* cenderung memiliki distribusi miring dengan keberadaan outlier. Outlier diidentifikasi menggunakan metode Z-Score dengan ambang $|Z| < 3$, mengacu pada pendekatan deteksi outlier dalam data ilmiah oleh Chandola [10]. Setelah eksplorasi, dilakukan pembersihan data dengan menghapus kolom yang tidak relevan dan memperbaiki nilai kosong. Variabel kategorikal *Area* dan *Item* diubah menjadi numerik menggunakan *Label Encoding* pada model awal, kemudian diganti menjadi *One-Hot Encoding* untuk meningkatkan representasi kategori. Fitur numerik distandarisasi menggunakan *StandardScaler*, terutama untuk memastikan stabilitas pembelajaran pada model *Gradient Boosting*. Dataset kemudian dibagi menjadi data latih (80%) dan data uji (20%).

3.2. Pembangunan Model dan Optimasi Parameter

Tiga algoritma regresi digunakan, yaitu Decision Tree Regressor, Random Forest Regressor, dan Gradient Boosting Regressor. Ketiganya dipilih karena efektif dalam menangani hubungan non-linear, bekerja baik pada data heterogen, dan banyak digunakan dalam penelitian prediksi hasil pertanian. Setelah pelatihan awal, dilakukan optimasi menggunakan *GridSearchCV* dengan *5-fold cross-validation* untuk meningkatkan performa dan mengurangi risiko *overfitting*. Untuk memperjelas ruang pencarian parameter, Tabel 1 berikut merangkum detail hyperparameter yang digunakan:

Tabel 1. Detail Hyperparameter

Model	Hyperparameter	Nilai
Decision Tree	max_depth	3, 5, 10, None
	min_samples_split	2, 5, 10
	min_samples_leaf	1, 2, 4

Random Forest	n_estimators max_depth min_samples_split min_samples_leaf	100, 200, 300 5, 10, None 2, 5, 10 1, 2, 4
Gradient Boosting	learning_rate n_estimators max_depth	0.01, 0.05, 0.1 100, 200, 300 3, 5

Tabel 1. Perbandingan Performa Model Regresi Machine Learning

Pemilihan hyperparameter tersebut didasarkan pada pengaruhnya terhadap kompleksitas dan kemampuan generalisasi model. Misalnya, peningkatan nilai *max_depth* memperbesar kompleksitas model namun dapat meningkatkan risiko *overfitting*, sedangkan peningkatan *n_estimators* pada model ensemble dapat meningkatkan stabilitas prediksi.

3.3. Hasil Evaluasi Model

Model dengan konfigurasi terbaik kemudian dievaluasi menggunakan empat metrik, yaitu Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan Coefficient of Determination (R^2). Keempat metrik ini memberikan gambaran menyeluruh tentang akurasi prediksi dan kesalahan model. Visualisasi tambahan seperti grafik *Predicted vs Actual*, *residual plot*, *feature importance*, serta *learning curve* digunakan untuk memperkuat interpretasi performa model dan memahami perilaku prediksinya secara lebih mendalam. Dengan serangkaian tahapan tersebut, metodologi penelitian ini memastikan bahwa proses analisis dilakukan secara sistematis, terukur, dan sesuai standar penelitian berbasis *machine learning*.

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil evaluasi tiga model regresi—*Decision Tree Regressor*, *Random Forest Regressor*, dan *Gradient Boosting Regressor*—setelah melalui tahapan preprocessing, penghapusan outlier, transformasi fitur, serta optimasi hyperparameter menggunakan *GridSearchCV*. Evaluasi dilakukan menggunakan data uji dengan empat metrik utama, yaitu MAE, MSE, RMSE, dan R^2 . Selain itu, disertakan visualisasi prediksi, analisis residual, *feature importance*, *learning curve*, serta pembahasan mendalam yang dikaitkan dengan temuan penelitian terdahulu dan interpretasi agronomi.

4.1. Hasil Evaluasi Model

Hasil evaluasi kuantitatif menunjukkan variasi performa yang cukup signifikan antara ketiga algoritma dapat dilihat pada tabel 2.

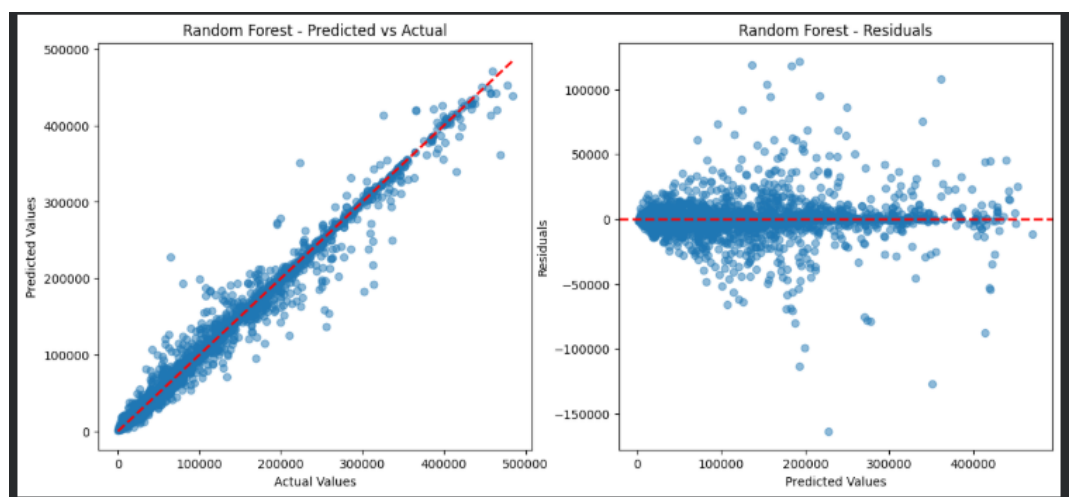
Tabel 2. Perbandingan Performa Model Regresi Machine Learning

Model	MAE	MSE	RMSE	R ²
Decision Tree Regressor	4.984.31	7.903.220.51	2.811.29	0.9742
Random Forest Regressor	3.053.90	3.264.620.11	1.806.83	0.9860
Gradient Boosting Regressor	5.142.76	8.605.511.23	2.933.39	0.9720

Hasil menunjukkan bahwa Random Forest Regressor menjadi model dengan performa terbaik, ditandai oleh nilai MAE dan RMSE terendah, serta nilai R² tertinggi mendekati 0,99. *Decision Tree Regressor* menunjukkan performa yang baik tetapi masih kalah stabil dibandingkan model ensemble. Sementara itu, *Gradient Boosting Regressor* menghasilkan performa terendah pada dataset ini.

4.2. Visualisasi Prediksi dan Residual

Untuk memperkuat pemahaman terhadap performa model, Gambar 2 menyajikan *Predicted vs Actual Plot* serta *Residual Plot* milik Random Forest.



Gambar 2. Plot Predicted vs Actual Random Forest

Plot *Predicted vs Actual* menunjukkan bahwa sebagian besar titik prediksi berada sangat dekat dengan garis diagonal, menandakan kesalahan prediksi yang relatif kecil. Nilai prediksi mengikuti pola nilai aktual secara konsisten, sehingga mendukung nilai R²

sebesar 0.9860. Pada *residual plot*, titik residual tersebar secara acak di sekitar garis horizontal nol tanpa pola tertentu. Pola ini mengindikasikan bahwa model tidak mengalami gejala heteroskedastisitas dan mampu melakukan generalisasi dengan baik pada berbagai rentang nilai hasil panen.

4.3. Pembahasan Performa Model

Keunggulan Random Forest Regressor pada penelitian ini dapat dijelaskan oleh beberapa faktor. Pertama, metode ensemble yang mengombinasikan banyak pohon keputusan memungkinkan model menangkap pola hubungan non-linear dan interaksi antar fitur secara lebih baik dibandingkan model tunggal. Dataset penelitian yang bersifat heterogen—terdiri dari 174 negara dan 52 komoditas—menuntut model yang mampu menyeimbangkan variasi antar kategori. Random Forest memiliki mekanisme *bootstrap aggregation* dan *random feature selection* yang membuatnya lebih tahan terhadap noise, ketidakseimbangan kategori, serta variasi ekstrem pada fitur numerik. Hal ini selaras dengan temuan Fitri & Nugraha [7] dan Hutahaean & Yusup [8], yang menyatakan bahwa Random Forest cenderung menjadi algoritma paling stabil untuk data agrikultur multinegara.

Pada sisi lain, *Decision Tree Regressor* hanya membangun satu struktur pohon sehingga rentan terhadap overfitting, terutama ketika data memiliki distribusi kompleks dan rentang nilai yang luas. Model cenderung terlalu mengikuti pola data latih, sehingga fleksibilitasnya dalam menghadapi variasi baru pada data uji menjadi terbatas. Meskipun demikian, nilai R^2 yang mendekati 0,97 menunjukkan bahwa Decision Tree tetap dapat menangkap pola dasar hubungan antar variabel.

Sementara itu, Gradient Boosting Regressor berada pada posisi terendah dalam evaluasi. Pada banyak studi terkontrol, boosting umumnya memberikan performa tinggi karena kemampuannya mengurangi error secara bertahap. Namun, pada dataset penelitian ini yang sangat heterogen, boosting menjadi lebih sensitif terhadap perbedaan skala, distribusi tidak merata, serta interaksi kategori yang kompleks. Meskipun standarisasi telah diterapkan, variabilitas data yang besar membuat algoritma kesulitan mempertahankan stabilitas. Fenomena ini juga dilaporkan oleh Himayanta & Wardhani [3], yang menemukan bahwa boosting menjadi kurang efektif ketika dataset memiliki distribusi yang sangat bervariasi.

4.4. Feature importance dan Analisis Model

Analisis *feature importance* dari Random Forest menunjukkan bahwa variabel *avg_rainfall*, *avg_temp*, *Area*, dan *Item* merupakan fitur yang paling berpengaruh dalam menentukan hasil panen. Dominansi fitur iklim sejalan dengan literatur agronomi yang menegaskan bahwa suhu dan curah hujan adalah faktor utama yang mempengaruhi

fisiologi tanaman serta produktivitas lahan. Variabel *item* dan *area* juga memberikan kontribusi besar karena masing-masing komoditas dan negara memiliki karakteristik agrikultur yang berbeda.

Kontribusi fitur yang seimbang menunjukkan bahwa Random Forest tidak hanya mengandalkan satu faktor dominan, tetapi mempertimbangkan interaksi multivariat antar variabel. Hal ini semakin menjelaskan mengapa model ensemble ini menunjukkan performa terbaik.

4.5. Hasil Learning Curve

Learning curve digunakan untuk menilai stabilitas performa model terhadap ukuran data latih. Grafik menunjukkan bahwa baik *training score* maupun *validation score* cenderung stabil pada nilai tinggi seiring peningkatan jumlah sampel, dengan jarak yang relatif kecil antara keduanya. Pola ini menunjukkan bahwa model tidak mengalami overfitting maupun underfitting secara signifikan. *Learning curve* yang stabil juga menegaskan bahwa dataset memiliki jumlah sampel yang memadai untuk melatih model ensemble.

4.6. Interpretasi Agronomi

Secara agronomis, hasil ini mendukung pemahaman bahwa prediksi hasil panen dipengaruhi secara kuat oleh kombinasi kondisi iklim dan karakteristik komoditas. Curah hujan dan suhu yang optimal berkorelasi dengan peningkatan produktivitas tanaman, sementara variasi antar negara dan komoditas mencerminkan perbedaan dalam praktik pertanian, kesuburan tanah, serta adaptasi tanaman. Konsistensi Random Forest dalam menangkap pola-pola tersebut menunjukkan bahwa pendekatan ensemble cocok untuk konteks prediksi hasil panen skala global yang melibatkan variabilitas tinggi.

4.7. Ringkasan

Hasil analisis menunjukkan bahwa Random Forest memberikan performa paling stabil dan akurat, didukung oleh visualisasi prediksi, residual, feature importance, dan learning curve. Temuan penelitian ini selaras dengan sebagian besar literatur agrikultur yang menekankan efektivitas model ensemble dalam menangani hubungan non-linear dan data multikategori. Sementara itu, Gradient Boosting menunjukkan performa lebih rendah akibat sensitivitas terhadap distribusi data yang tidak merata. Dengan demikian, Random Forest menjadi model paling direkomendasikan untuk prediksi hasil panen pada dataset global yang kompleks.

5. KESIMPULAN

Penelitian ini membandingkan tiga algoritma regresi—*Decision Tree Regressor*, *Random Forest Regressor*, dan *Gradient Boosting Regressor*—dalam memprediksi hasil panen menggunakan dataset FAOSTAT. Setelah melalui preprocessing, penanganan outlier, transformasi fitur, serta optimasi dengan *GridSearchCV*, diperoleh bahwa *Random Forest Regressor* menjadi model paling unggul dengan nilai MAE dan RMSE terendah serta R^2 tertinggi (0.9860). Stabilitas model ini dipengaruhi oleh mekanisme ensemble yang efektif menangani data agrikultur yang heterogen. *Decision Tree Regressor* memberikan hasil yang cukup baik namun cenderung overfitting, sedangkan *Gradient Boosting Regressor* menunjukkan performa terendah akibat sensitivitasnya terhadap distribusi data yang tidak merata. Analisis *feature importance* menunjukkan bahwa curah hujan, suhu, serta karakteristik negara dan komoditas merupakan faktor utama yang memengaruhi hasil panen, selaras dengan kajian agronomi. Penelitian ini memperkuat literatur yang menyatakan bahwa metode ensemble berbasis pohon cocok untuk prediksi hasil panen berskala besar. Temuan ini bukan merupakan kebaruan algoritmik, namun memberikan evaluasi komparatif yang terstruktur dan relevan sebagai dasar pengembangan sistem pendukung keputusan di bidang pertanian.

DAFTAR PUSTAKA

- [1] Frira Sesilia, Viktor Handrianus Pranatawijaya, and Ressa Priskila, "Machine Learning untuk Memprediksi Jumlah Penjualan, Stok dan Jumlah Tanam Hasil Pertanian Hidroponik," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 4, no. 1, pp. 222–233, 2024, doi: 10.24002/konstelasi.v4i1.9055.
- [2] P. H. Putra, J. Julham, N. Nurlinda, and I. Dhitisari, "Inovasi Smart Farming Optimalisasi Bawang Merah Hidroponik Berbasis Iot Dan Machine Learning," *J. Sci. Soc. Res.*, vol. 7, no. 4, p. 1788, 2024, doi: 10.54314/jssr.v7i4.2330.
- [3] N. Afrilia S, F. Frazna Az-Zahra, and P. Prajoko, "Prediksi Hasil Panen Wortel Menggunakan Algoritma Regresi Linear Berganda," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 5, pp. 10255–10262, 2024, doi: 10.36040/jati.v8i5.10954.
- [4] M. F. Anggarda, I. Kustiawan, D. R. Nurjanah, and N. F. A. Hakim, "Pengembangan Sistem Prediksi Waktu Penyiraman Optimal pada Perkebunan: Pendekatan Machine Learning untuk Peningkatan Produktivitas Pertanian," *J. Budid. Pertan.*, vol. 19, no. 2, pp. 124–136, 2023, doi: 10.30598/jbdp.2023.19.2.124.
- [5] T. Nizami, M. A. Mustaqim, and W. Ariannor, "Analisis kinerja model machine learning dalam prediksi gagal panen gabah," *Progresif J. Ilm. Komput.*, vol. 21, no. 1, pp. 184–192, 2025, [Online]. Available: <https://ojs.stmik-banjarbaru.ac.id/index.php/progresif/article/view/2501>
- [6] Salahuddin, M. Khadafi, Huzeini, and M. Davi, "Rancang Bangun Aplikasi Machine

- Learning Prediksi Hasil Panen Buah Pinang (ArecaCatechu) Menggunakan Metode Regresi Linier Berganda,” *Proceeding Semin. Nas. Politek. Negeri Lhokseumawe*, vol. 6, no. 1, pp. 181–186, 2022.
- [7] E. Fitri and S. N. Nugraha, “Optimasi Kinerja Linear Regression, Random Forest Regression Dan Multilayer Perceptron Pada Prediksi Hasil Panen,” *INTI Nusa Mandiri*, vol. 18, no. 2, pp. 210–217, 2024, doi: 10.33480/inti.v18i2.5269.
- [8] K. L. Himayanta and D. F. Wardhani, “PREDIKSI HASIL PANEN PADI DENGAN MACHINE LEARNING,” vol. 13, no. 1, pp. 1–14, 2023.
- [9] P. Jonatan Hutahaeen, Dadang Yusup, “Perbandingan Metode Linear Regression , Random Forest & K-Nearest Neighbor Untuk Prediksi Produksi,” vol. 8, no. 3, pp. 3895–3900, 2024.
- [10] Z. A. N. A. Fareza, I. Cholissodin, and L. Muflikhah, “Prediksi Hasil Panen Tanaman Biofarmaka di Indonesia dengan Menggunakan Metode Extreme Learning Machine,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 11, p. 5331, 2022.