

Pemodelan Produksi Selada Berbasis Faktor Lingkungan dan Metode Budidaya dengan Pendekatan Machine Learning

Rizki Hesnananda¹, Muhamad Rifai² Dana Santoso Saroso³

¹Jurusan Teknologi Informasi, Universitas Siber Indonesia, Jakarta

²Jurusan Teknik Informatika, Fakultas Teknik, Universitas Krisnadwipayana, Jakarta

³Jurusan Bisnis Digital, Universitas Siber Indonesia, Jakarta

Abstrak. Penelitian ini bertujuan membangun model prediksi hasil panen selada menggunakan variabel lingkungan dan budidaya yang meliputi suhu, curah hujan, tingkat keasaman tanah (pH), kelembaban tanah, jumlah pemupukan, luas lahan, jenis pupuk, serta metode irigasi. Dataset terdiri dari 1.000 entri tanpa nilai hilang dan mencakup kombinasi fitur numerik dan kategorikal. Analisis awal menunjukkan bahwa kelembaban tanah, jumlah pemupukan, dan curah hujan memiliki korelasi positif yang lebih kuat terhadap hasil panen dibandingkan variabel lain. Proses persiapan data meliputi pembersihan data, standarisasi fitur numerik, pengodean fitur kategorikal, serta pembagian data menggunakan skema 80:20. Tiga algoritma diterapkan, yaitu Linear Regression, Random Forest Regressor, dan XGBoost Regressor. Evaluasi menggunakan MAE, RMSE, dan R^2 memperlihatkan bahwa XGBoost memberikan performa terbaik dengan R^2 sebesar 0,96, diikuti oleh Random Forest ($R^2 = 0,86$), sementara Linear Regression menunjukkan performa paling rendah ($R^2 = 0,42$). Hasil penelitian menegaskan bahwa metode ensemble, khususnya XGBoost, lebih efektif dalam menangkap hubungan kompleks antar variabel pada konteks prediksi hasil panen selada. Temuan ini berpotensi mendukung pengambilan keputusan bagi petani dalam perencanaan produksi secara lebih akurat.

Kata kunci—*Prediksi hasil panen; Selada; Machine learning; XGBoost; Random Forest.*

1. PENDAHULUAN

Pertanian modern menuntut penggunaan data sebagai dasar pengambilan keputusan. Variabel lingkungan seperti suhu, curah hujan, kelembaban tanah, dan kondisi kimia tanah berperan langsung terhadap produktivitas tanaman hortikultura, termasuk selada [1], [2]. Ketidakpastian faktor iklim dan variasi praktik budidaya menyebabkan hasil panen sulit diprediksi secara konsisten [3][4]. Dalam kondisi seperti ini, pendekatan berbasis data diperlukan untuk memperoleh estimasi yang lebih terukur mengenai potensi hasil produksi [5], [6], [7].

Model prediksi hasil panen memberikan fungsi strategis bagi petani maupun pelaku usaha agrikultur. Informasi prediktif memungkinkan perencanaan pemupukan, penjadwalan irigasi, distribusi tenaga kerja, hingga penentuan volume penjualan [8], [9], [10]. Ketika praktik budidaya dilakukan di lahan kecil hingga menengah dengan sumber daya terbatas, ketepatan perhitungan menjadi faktor penentu efisiensi biaya dan stabilitas produksi [11], [12], [13]. Oleh sebab itu, penggunaan metode komputasional menjadi relevan untuk mengatasi variabilitas kondisi lapangan.

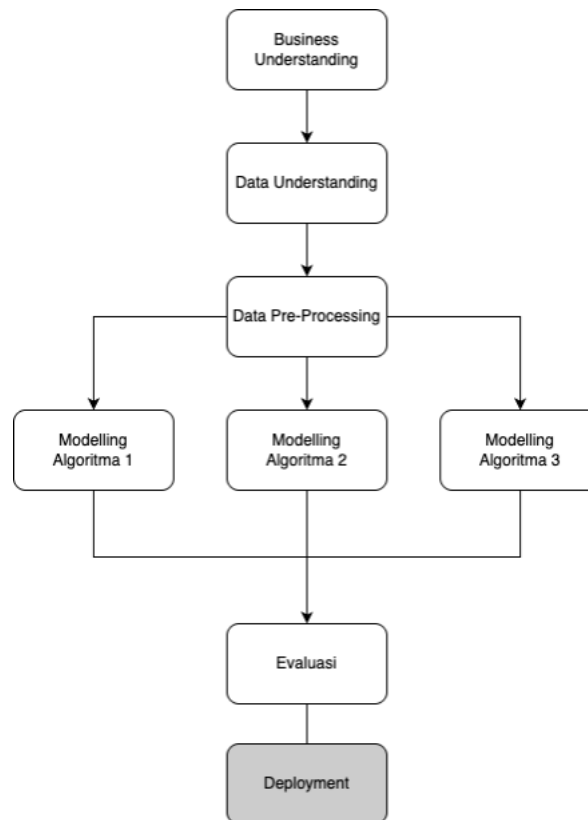
Perkembangan pembelajaran mesin menawarkan metode yang lebih kuat dalam memodelkan hubungan kompleks antar variabel pertanian [14], [15], [16]. Algoritma seperti Linear Regression memberikan baseline interpretatif, sementara metode ensemble seperti Random Forest dan XGBoost mampu menangani hubungan non-linear serta interaksi antar fitur [17], [18]. Kinerja algoritma-algoritma tersebut menjadikannya alat penting dalam analisis prediktif sektor pertanian, khususnya ketika data mencakup kombinasi fitur numerik dan kategorikal [19], [20].

Penelitian ini memanfaatkan dataset hasil panen selada yang mencakup 1.000 observasi dengan sembilan variabel utama terkait kondisi lingkungan, teknik budidaya, dan hasil panen. Analisis pendahuluan menunjukkan bahwa kelembaban tanah, curah hujan, dan jumlah pemupukan memiliki korelasi yang lebih dominan terhadap hasil panen dibandingkan variabel lain. Struktur data yang bersih dan lengkap memungkinkan penerapan langsung proses transformasi, pelatihan model, dan evaluasi performa.

Penelitian ini bertujuan membandingkan performa tiga algoritma prediksi—Linear Regression, Random Forest Regressor, dan XGBoost Regressor—untuk mengidentifikasi metode yang paling efektif dalam memprediksi hasil panen selada. Evaluasi dilakukan menggunakan metrik MAE, RMSE, dan R^2 untuk memastikan penilaian menyeluruh. Hasil penelitian memberikan dasar empiris bagi pengembangan sistem pendukung keputusan di sektor pertanian dan membuka peluang implementasi model prediktif pada komoditas hortikultura lainnya.

2. METODE

Penelitian ini menggunakan kerangka kerja CRISP-DM [21], [22] yang terdiri dari enam tahap utama seperti ditampilkakan pada Gambar 1 berikut:



Gambar 1. Alur penelitian CRISP-DM

Pada tahap *Business Understanding*, penelitian difokuskan pada perumusan tujuan untuk membangun model prediksi hasil panen selada berdasarkan variabel lingkungan dan praktik budidaya. Tahap ini mencakup identifikasi kebutuhan pengguna, ruang lingkup penelitian, serta tujuan akhir berupa model prediktif yang dapat meningkatkan akurasi estimasi produksi.

Tahap *Data Understanding* direncanakan melalui eksplorasi awal dataset yang memuat variabel suhu, curah hujan, pH tanah, kelembaban tanah, jumlah pemupukan, luas lahan, jenis pupuk, metode irigasi, dan hasil panen. Aktivitas ini meliputi pengenalan struktur data, pemeriksaan kelengkapan, identifikasi tipe variabel, serta pengamatan pola awal antar fitur untuk memahami karakteristik dataset sebelum pemrosesan lebih lanjut.

Pada tahap *Data Preparation*, data akan dipersiapkan agar memenuhi kebutuhan pemodelan. Proses ini meliputi pembersihan data, transformasi fitur numerik, pengodean fitur kategorikal, serta pembagian data menjadi set pelatihan dan pengujian. Tahap ini dirancang untuk memastikan dataset berada dalam kondisi yang konsisten dan layak digunakan oleh algoritma pembelajaran mesin.

Tahap *Modeling* direncanakan dengan menerapkan beberapa algoritma regresi, seperti Linear Regression, Random Forest Regressor, dan XGBoost Regressor. Setiap algoritma dipilih untuk mewakili model linear dan non-linear sehingga memungkinkan perbandingan performa antar pendekatan. Penyesuaian parameter model juga direncanakan apabila dibutuhkan untuk meningkatkan akurasi prediksi.

Tahap *Evaluation* akan dilakukan untuk menilai kualitas model menggunakan metrik MAE, RMSE, dan R^2 . Evaluasi bertujuan memastikan bahwa model yang dihasilkan memenuhi tujuan penelitian dan mampu memberikan prediksi yang memadai. Pada penelitian ini, proses berhenti pada tahap evaluasi dan tidak mencakup tahap *Deployment*, sehingga model tidak diimplementasikan pada sistem produksi atau aplikasi lapangan.

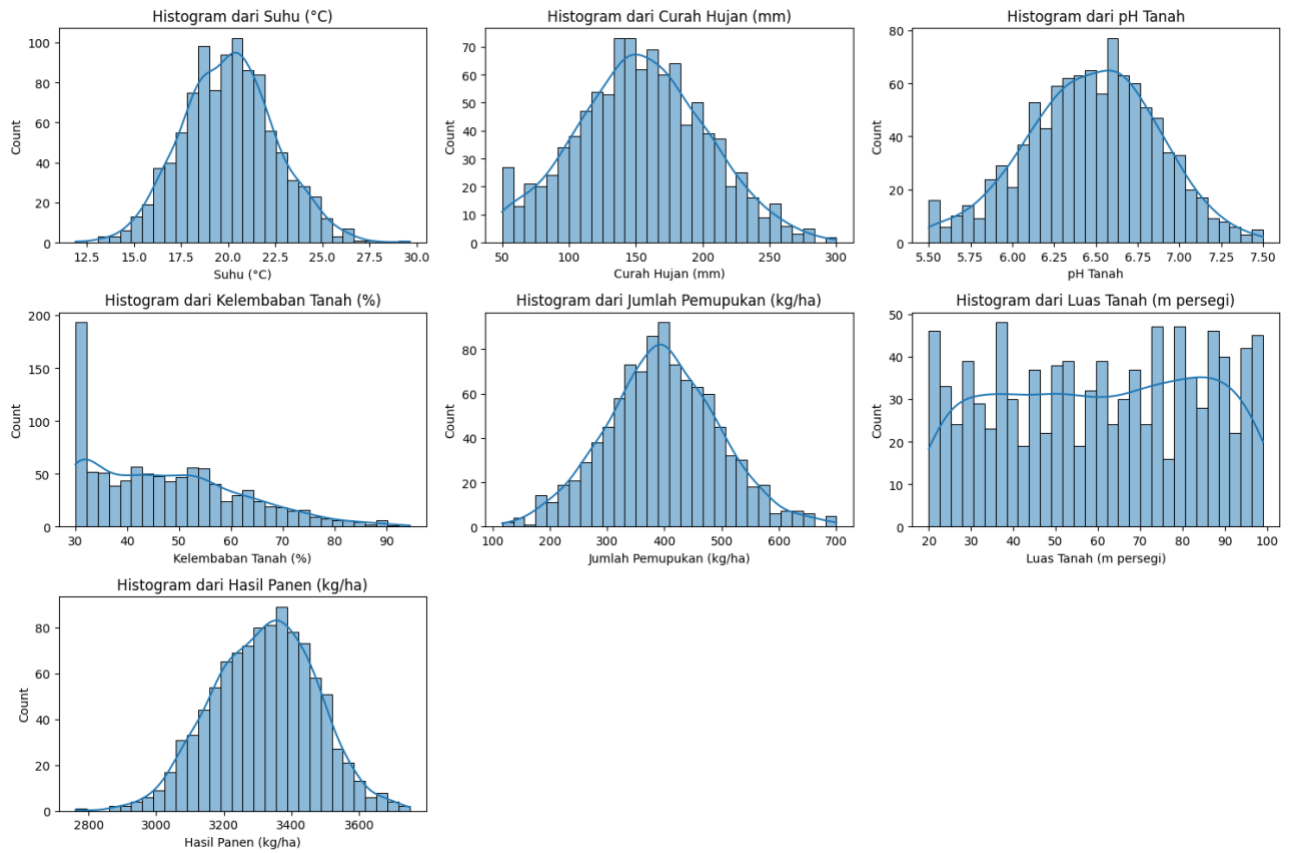
3. HASIL DAN PEMBAHASAN

a. *Business Understanding*

Permasalahan utama dalam penelitian ini adalah kebutuhan untuk memperkirakan jumlah hasil panen selada berdasarkan sejumlah faktor lingkungan dan praktik budidaya. Variabel yang digunakan meliputi suhu, curah hujan, tingkat keasaman tanah, kelembaban tanah, jumlah pupuk, luas lahan, jenis pupuk, serta metode irigasi. Dalam praktik pertanian, hasil panen sering berfluktuasi karena dipengaruhi perubahan cuaca dan kondisi tanah, sehingga menjadi sulit diprediksi secara konsisten. Model prediksi dibutuhkan agar petani memiliki gambaran kuantitatif mengenai potensi hasil panen, sehingga perencanaan penggunaan pupuk, air, biaya operasional, dan strategi penjualan dapat dilakukan lebih efisien. Fokus prediksi diarahkan pada jumlah panen selada sebagai variabel target karena informasi ini dapat dimanfaatkan langsung untuk pengambilan keputusan budidaya maupun perencanaan usaha.

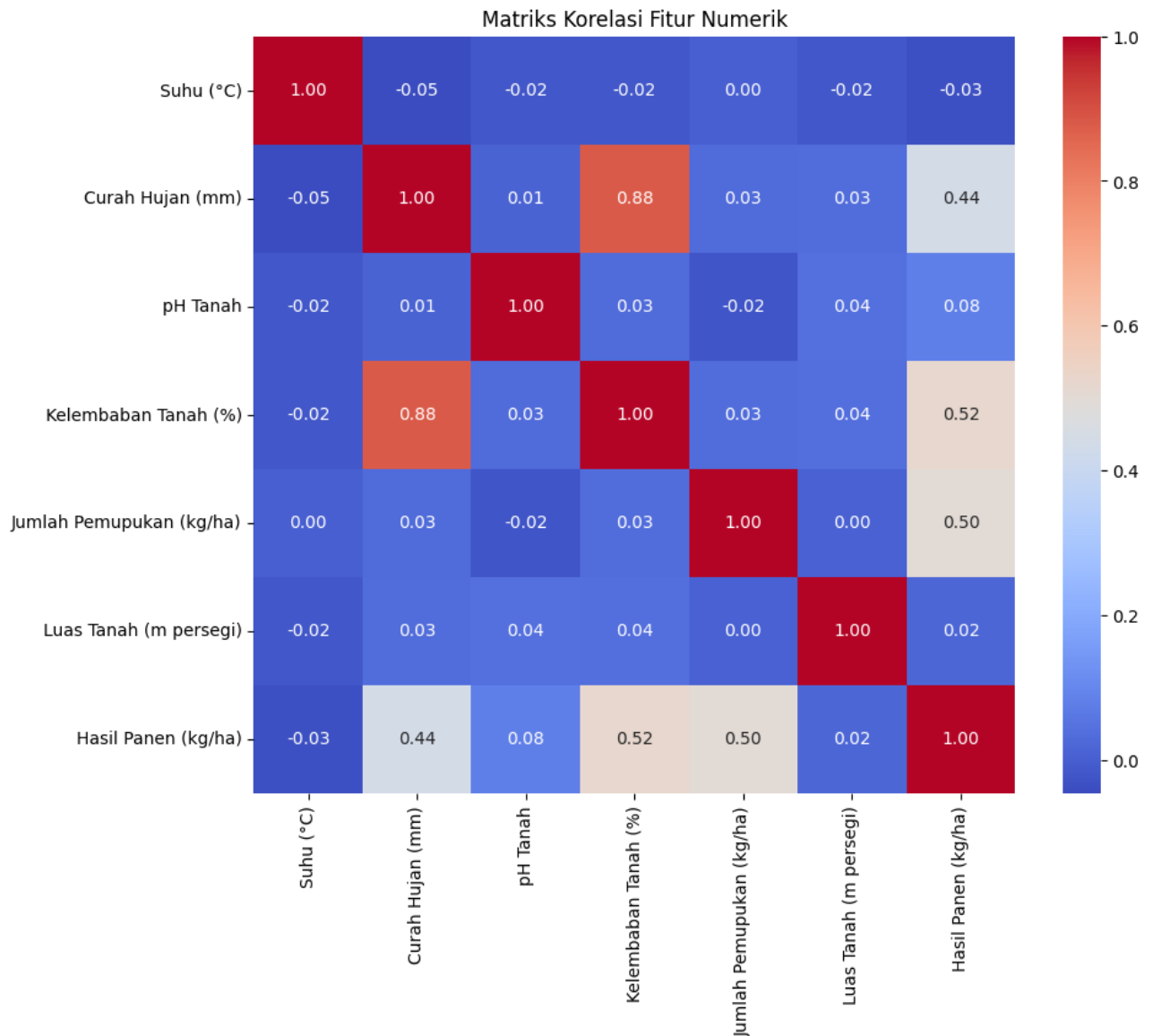
b. *Data Understanding*

Dataset penelitian mencakup 1.000 observasi dengan sembilan fitur yang relevan untuk memprediksi hasil panen selada. Tidak terdapat nilai kosong sehingga seluruh data siap dianalisis.



Gambar 2. Histogram Data Selada

Sebaran data setiap variabel dapat dilihat pada Gambar 2 dalam bentuk histogram. Empat fitur memiliki tipe data float, tiga di antaranya adalah integer, dan dua fitur berupa kategori. Gambaran statistik awal menunjukkan rentang variasi yang cukup lebar pada sejumlah variabel seperti suhu, curah hujan, dan kelembaban tanah.



Gambar 3. Heatmap Matriks Korelasi

Pada Gambar 3, ditampilkan hasil korelasi memperlihatkan bahwa kelembaban tanah memiliki hubungan paling kuat dengan curah hujan, sekaligus menunjukkan hubungan positif terhadap hasil panen. Jumlah pemupukan dan curah hujan juga memberikan kontribusi korelasional, sementara suhu, pH tanah, dan luas lahan terlihat memiliki pengaruh langsung yang lebih rendah dalam konteks dataset ini.

c. Data Preparation

Tahap persiapan data dilakukan untuk memastikan seluruh variabel berada dalam format yang sesuai untuk proses pemodelan. Data dibersihkan dari kemungkinan nilai tidak lengkap, kemudian fitur diklasifikasikan menjadi numerik dan kategorikal agar dapat diproses menggunakan teknik transformasi yang sesuai. Fitur numerik direncanakan untuk dinormalisasi agar perbedaan skala antar variabel tidak mempengaruhi pelatihan model, sedangkan fitur kategorikal akan dikodekan menjadi representasi numerik menggunakan metode *one-hot encoding*. Setelah itu, dataset dibagi menjadi data pelatihan dan pengujian dengan proporsi 80:20 untuk memberikan dasar evaluasi terhadap kemampuan generalisasi model.

d. Modelling

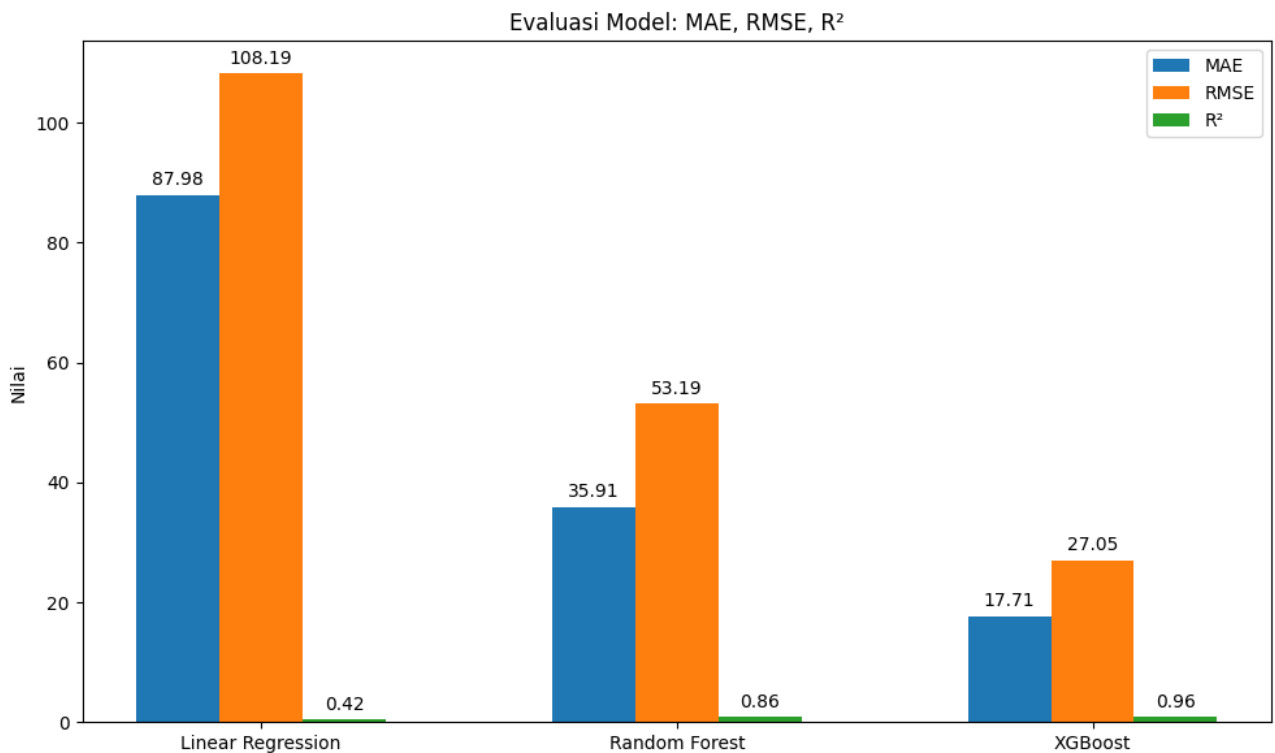
Tiga algoritma regresi dipilih untuk membangun model prediksi, yaitu Linear Regression sebagai baseline, serta Random Forest dan XGBoost sebagai model berbasis ensemble yang mampu menangani hubungan non-linear. Setiap model dilatih menggunakan data yang telah diproses. Untuk model yang lebih kompleks seperti Random Forest dan XGBoost, direncanakan penggunaan penyesuaian hyperparameter melalui teknik pencarian terstruktur guna memperoleh konfigurasi yang memberikan kinerja terbaik. Proses ini memungkinkan perbandingan yang lebih adil antar model.

e. Evaluation

Tabel 1. Perbandingan Hasil Evaluasi

	Linear Regression	Random Forest	XGBoost
MAE	87.98	35.91	17.71
RMSE	108.19	53.19	27.05
R-Squared	0.42	0.86	0.96

Tabel 1 menampilkan evaluasi model dilakukan dengan menggunakan metrik MAE, RMSE, dan R² untuk menilai tingkat kesalahan dan kemampuan model dalam menjelaskan variasi data. Linear Regression diperkirakan menghasilkan performa dasar, sementara model ensemble diharapkan menunjukkan akurasi lebih tinggi.



Gambar 4. Evaluasi Tiap Model

Pada Gambar 4, ditampilkan hasil evaluasi digunakan untuk menentukan model yang paling layak diterapkan pada konteks prediksi hasil panen. Penelitian ini berhenti pada tahap evaluasi dan tidak melanjutkan ke proses deployment, sehingga model tidak diterapkan ke sistem produksi atau aplikasi pengguna.

KESIMPULAN

Penelitian ini menunjukkan bahwa pendekatan pembelajaran mesin berpotensi memberikan estimasi hasil panen selada yang lebih akurat dibandingkan metode perhitungan tradisional. Dengan memanfaatkan variabel lingkungan dan praktik budidaya sebagai fitur prediktif, model regresi mampu menangkap hubungan antara kondisi lapangan dan produktivitas tanaman. Pendekatan ini memberikan dasar sistematis bagi pengembangan alat bantu prediksi yang dapat meningkatkan efisiensi perencanaan pertanian.

Hasil evaluasi membuktikan bahwa algoritma berbasis ensemble, seperti Random Forest dan terutama XGBoost, memberikan performa prediksi yang lebih unggul dibandingkan model linear. Temuan ini mengindikasikan bahwa pola hubungan antar variabel dalam konteks pertanian bersifat kompleks dan lebih tepat dimodelkan menggunakan metode non-linear. Kesimpulan ini memberikan arah bagi penelitian lanjutan, baik dalam perluasan variabel, peningkatan kualitas data, maupun pengembangan sistem pendukung keputusan berbasis machine learning.

DAFTAR PUSTAKA

- [1] S. Sisriana, S. Suryani, S. S.-J. I. Respati, and undefined 2021, “Pengaruh berbagai media tanam terhadap pertumbuhan dan kadar pigmen microgreens selada,” *ejournal.urindo.ac.id*, vol. 12, no. Desember, pp. 1411–7126, 2021, Accessed: Nov. 27, 2025. [Online]. Available: <https://ejournal.urindo.ac.id/index.php/pertanian/article/view/1886>
- [2] A. Dewi, N. Lubis, S. S.-P. T. Media, and undefined 2023, “Budidaya Selada Organik Ramah Lingkungan,” *tahtamedia.co.id*, Accessed: Nov. 27, 2025. [Online]. Available: <http://tahtamedia.co.id/index.php/issj/article/view/125>
- [3] E. Grigorieva, A. Livenets, and E. Stelmakh, “Adaptation of agriculture to climate change: A scoping review,” *Climate*, vol. 11, no. 10, p. 202, 2023.
- [4] D. L. T. Anh, N. T. Anh, and A. A. Chandio, “Climate change and its impacts on Vietnam agriculture: A macroeconomic perspective,” *Ecol Inform*, vol. 74, p. 101960, 2023.
- [5] N. Afrilia, F. Az-Zahra, P. P.-J. (Jurnal M. Teknik, and undefined 2024, “Prediksi hasil panen wortel menggunakan algoritma regresi linear berganda,” *ejournal.itn.ac.id*, vol. 8, no. 5, 2024, Accessed: Nov. 27, 2025. [Online]. Available: <https://www.ejournal.itn.ac.id/jati/article/view/10954>
- [6] A. Maulana, M. Martanto, I. A.-J. (Jurnal M. Teknik, and undefined 2023, “Prediksi Hasil Produksi Panen Bawang Merah Menggunakan Metode Regresi Linier Sederhana,” *ejournal.itn.ac.id*, vol. 7, no. 4, 2023, Accessed: Nov. 27, 2025. [Online]. Available: <https://www.ejournal.itn.ac.id/jati/article/view/7281>
- [7] B. A.-J. F. E. E. Listrik and undefined 2022, “Implementasi Metode Double Exponential Smoothing Untuk Prediksi Hasil Panen Sayuran Kentang,” *elektroda.uho.ac.id*, Accessed: Nov. 27, 2025. [Online]. Available: <https://elektroda.uho.ac.id/index.php/journal/article/view/9>
- [8] I. Negara, ... I. N.-... T. I. D., and undefined 2023, “Prediksi Hasil Panen Padi Di Kabupaten Jembrana Menggunakan Metode Naive Bayes Classifier,” *jurnal.undhirabali.ac.id*, Accessed: Nov. 27, 2025. [Online]. Available: <https://jurnal.undhirabali.ac.id/index.php/jutik/article/view/2501>
- [9] Z. Fareza, ... I. C.-... T. I. dan, and undefined 2022, “Prediksi hasil panen tanaman biofarmaka di Indonesia dengan menggunakan metode extreme learning machine,” *j-ptiik.ub.ac.id*, Accessed: Nov. 27, 2025. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11867>
- [10] D. Nuraini, D. Violina, ... D. A.-J. J., and undefined 2025, “Prediksi Hasil Panen Padi dengan Metode Multiple Linear Regression dan Particle Swarm Optimization untuk Meningkatkan Produksi Padi di Madura,” *ejournal.uniramalang.ac.id*, Accessed: Nov. 27, 2025. [Online]. Available: <https://ejournal.uniramalang.ac.id/jusifor/article/view/5857>
- [11] C. H.-S. L. J. I. Indonesia and undefined 2022, “Sistem hidroponik menggunakan nutrient film technique untuk produksi dan hasil tanaman selada (*lactuca sativa* l.),” *jurnal.syntaxliterate.co.id*, Accessed: Nov. 27, 2025. [Online]. Available: <https://jurnal.syntaxliterate.co.id/index.php/syntax-literate/article/view/11983>
- [12] I. Lestari, A. Rahayu, Y. M.-J. Agronida, and undefined 2022, “Pertumbuhan Dan Produksi Tanaman Selada (*Lactuca sativa* L.) pada berbagai media tanam dan konsentrasi nutrisi pada sistem hidroponik

- Nutrient Film Technique,” *ojs.unida.ac.id*, vol. 8, no. 1, p. 31, 2022, Accessed: Nov. 27, 2025. [Online]. Available: <https://ojs.unida.ac.id/JAG/article/view/5625>
- [13] A. L.-B. J. I. Pertanian and undefined 2021, “Pengaruh konsentrasi dan interval pemberian POC urin kelinci terhadap pertumbuhan dan produksi tanaman selada (*Lactuca sativa* L.),” *jurnal.unikal.ac.id*, Accessed: Nov. 27, 2025. [Online]. Available: <http://jurnal.unikal.ac.id/index.php/biofarm/article/view/1610/1085>
- [14] N. Hadi, J. B.-C. J. of C. Science, and undefined 2024, “Implementasi Machine Learning Untuk Prediksi Harga Rumah Menggunakan Algoritma Random Forest,” *journal.untar.ac.id*, vol. 8, no. 1, pp. 50–61, 2024, Accessed: Nov. 26, 2025. [Online]. Available: <https://journal.untar.ac.id/index.php/computatio/article/view/15173>
- [15] R. Hesananda, *Algoritma Klasifikasi Bibit Terbaik untuk Tanaman Keladi Tikus*. Penerbit NEM, 2021.
- [16] R. Hesananda and D. F. Racma, “IMPLEMENTASI GOOGLE LOOKER STUDIO UNTUK ANALISIS TREN DAN VISUALISASI DATA (STUDI KASUS: PRODUKSI PADI PULAU SUMATERA),” *Innotech: Jurnal Ilmu Komputer, Sistem Informasi dan Teknologi Informasi*, vol. 1, no. 2, pp. 62–74, 2024.
- [17] X. Li *et al.*, “Exploring interactive and nonlinear effects of key factors on intercity travel mode choice using XGBoost,” *Elsevier*, Accessed: Nov. 27, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0143622824000699>
- [18] S. Fatima, A. Hussain, S. Amir, ... M. A.-P. journal of, and undefined 2023, “Xgboost and random forest algorithms: an in depth analysis,” *paas-pk.org*, Accessed: Nov. 27, 2025. [Online]. Available: <https://paas-pk.org/index.php/pjosr/article/view/1876>
- [19] R. Faizal, ... A. A.-... (Computer S. and, and undefined 2025, “Perbandingan Random Forest Regressor Dan Decision Tree Regressor Untuk Prediksi Hasil Panen,” *ejurnal.umri.ac.id*, vol. 6, no. 2, pp. 247–253, 2025, doi: 10.37859/coscitech.v6i2.9966.
- [20] A. Nugroho, E. R.-Techno. com, and undefined 2023, “Penerapan Metode Oversampling SMOTE Pada Algoritma Random Forest Untuk Prediksi Kebangkrutan Perusahaan.,” *search.ebscohost.com*, Accessed: Nov. 26, 2025. [Online]. Available: <https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=14122693&AN=164068345&h=DscQNqI8urjYQhwBgRJn01b8IKXnXur1VhCm9uw8PIUdj2ucwMG4rYznnY%2B5Lb4CrIEiBjD7P6FPQGBB623Q%3D%3D&crl=c>
- [21] C. Schröer, F. Kruse, J. G.-P. computer science, and undefined 2021, “A systematic literature review on applying CRISP-DM process model,” *Elsevier*, Accessed: Nov. 26, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050921002416>
- [22] J. S. Saltz, “CRISP-DM for data science: strengths, weaknesses and potential next steps,” in *2021 IEEE International Conference on Big Data (Big Data)*, IEEE, 2021, pp. 2337–2344.