



PREDIKSI PANEN KELAPA SAWIT MENGGUNAKAN METODE RANDOM FOREST REGRESSION

Friska Dwi Choirunisa, Menur Wahyu Pangestika, Sucipto

Prodi Teknik Informatika, Universitas Muhammadiyah Pontianak

Abstrak

Tanaman kelapa sawit termasuk dalam komoditas perkebunan unggulan yang memiliki kontribusi signifikan terhadap pertumbuhan ekonomi, sehingga dibutuhkan strategi perencanaan panen yang tepat guna mengoptimalkan pemanfaatan sumber daya yang tersedia. Kelompok Tani "ASRI" hingga saat ini belum memiliki sistem yang mampu memperkirakan hasil panen secara tepat, sehingga proses perencanaan masih bergantung pada cara konvensional. Penelitian ini bertujuan membangun sebuah sistem prediksi hasil panen kelapa sawit dengan menggunakan metode Random Forest Regression yang berlandaskan data historis panen. Variabel yang digunakan mencakup tahun panen, bulan panen, usia tanaman, jenis tanah, pemberian pupuk, kondisi musim, dan luas areal tanam. Alur penelitian meliputi pengumpulan data, tahap preprocessing yang terdiri dari pembersihan data, penanganan data pencilan, normalisasi, dan label encoding, kemudian dilanjutkan dengan pemisahan data menjadi data latih dan data uji, pelatihan model Random Forest Regression, serta evaluasi kinerja model menggunakan Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE). Berdasarkan hasil pengujian, model menghasilkan nilai MAE sebesar 869,75 kg/bulan dan RMSE sebesar 1111,89 kg/bulan, yang menunjukkan bahwa model mampu menghasilkan prediksi hasil panen dengan tingkat deviasi yang tergolong rendah. Model yang telah dikembangkan selanjutnya diimplementasikan dalam bentuk aplikasi berbasis Streamlit sebagai alat bantu dalam perencanaan panen dan proses pengambilan keputusan di Kelompok Tani "ASRI".

Kata Kunci: Kelapa Sawit, Prediksi Hasil Panen, Random Forest Regression, Machine Learning, Streamlit.

PENDAHULUAN

Kelapa sawit merupakan komoditas Perkebunan yang memiliki

peranan penting dalam mendukung pertumbuhan ekonomi Indonesia. Komoditas ini memberikan kontribusi

yang signifikan terhadap sektor perkebunan serta menjadi salah satu sumber pendapatan nasional. Produktivitas kelapa sawit dipengaruhi oleh berbagai faktor, seperti umur tanaman, jenis tanah, kondisi musim, pemupukan, dan luas lahan. Oleh karena itu, informasi mengenai perkiraan hasil panen sangat dibutuhkan untuk mendukung perencanaan produksi, pengelolaan sumber daya, serta pengambilan keputusan yang lebih efektif.

Kelompok Tani "ASRI" masih mengandalkan pengalaman dan perkiraan dalam menentukan hasil panen, sehingga ketepatan prediksi yang diperoleh belum optimal dan berpotensi memengaruhi proses perencanaan produksi. Ketidakakuratan prediksi dapat berdampak pada perencanaan distribusi, pengelolaan tenaga kerja, dan penyediaan sarana produksi. Seiring perkembangan teknologi, metode Machine Learning dapat dimanfaatkan untuk membantu memprediksi hasil panen berdasarkan data historis dan faktor-faktor yang memengaruhinya.

Salah satu algoritma Machine Learning yang banyak digunakan untuk prediksi adalah Random Forest Regression. Algoritma ini memiliki kemampuan dalam menangani hubungan non-linear antar variabel, tahan terhadap noise dan outlier, serta mampu menghasilkan performa prediksi yang baik pada berbagai kasus. Studi sebelumnya membuktikan bahwa metode Random Forest Regression mampu menghasilkan prediksi yang cukup akurat dalam kasus estimasi produksi padi di Jawa Barat, yang ditunjukkan dengan perolehan nilai koefisien determinasi (R^2) sebesar 0,96. Studi lain mengungkapkan bahwa metode Random Forest Regression menunjukkan performa yang lebih unggul dibandingkan dengan regresi

linear dalam menghasilkan prediksi berbasis data historis.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan sistem prediksi hasil panen kelapa sawit dengan menerapkan metode Random Forest Regression berdasarkan variabel tahun panen, bulan panen, umur tanaman, jenis tanah, pemupukan, musim, dan luas lahan. Evaluasi model dilakukan menggunakan metrik Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE). Hasil penelitian menunjukkan bahwa model yang dibangun memperoleh nilai MAE sebesar 869,75 kg/bulan dan RMSE sebesar 1111,89 kg/bulan sehingga dapat digunakan sebagai alat bantu dalam perencanaan panen dan pengambilan keputusan berbasis data.

METODE PENELITIAN

Penelitian ini difokuskan pada pengembangan sistem prediksi hasil panen kelapa sawit menggunakan metode Random Forest Regression berdasarkan data historis Kelompok Tani "ASRI". Tahapan penelitian meliputi identifikasi masalah, studi literatur, pengumpulan data, preprocessing data, perancangan dan implementasi model, pengujian, serta penarikan kesimpulan.

Data yang digunakan terdiri dari variabel tahun dan bulan panen, umur tanaman, jenis tanah, pemupukan, kondisi musim, luas lahan, serta hasil panen sebagai target. Data diperoleh dari catatan historis kelompok tani dan dilakukan proses pembersihan sebelum digunakan.

Pada tahap preprocessing dilakukan pembersihan data, transformasi variabel kategorikal menggunakan label encoding, serta pembagian data menjadi 80% data latih dan 20% data uji. Model Random Forest Regression dibangun dan dilatih menggunakan parameter tertentu

seperti jumlah pohon ($n_estimators$) dan kedalaman maksimum (max_depth).

Evaluasi model dilakukan menggunakan Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE), serta analisis feature importance untuk mengetahui variabel paling berpengaruh. Model kemudian diimplementasikan ke dalam aplikasi berbasis Streamlit untuk mendukung proses prediksi dan pengambilan keputusan secara otomatis.

HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan sebuah aplikasi prediksi hasil panen kelapa sawit berbasis web yang dikembangkan menggunakan framework Streamlit. Framework Streamlit digunakan karena mampu membangun aplikasi web interaktif berbasis Python dengan mudah serta mendukung integrasi langsung dengan model Machine Learning. Sistem prediksi yang dikembangkan menggunakan algoritma Random Forest Regression sebagai metode utama untuk memprediksi hasil panen kelapa sawit.

Langkah awal dalam penelitian ini diawali dengan pelaksanaan preprocessing data guna memastikan bahwa data yang digunakan dalam proses pemodelan memiliki kualitas yang baik dan layak untuk diolah lebih lanjut. Proses tersebut meliputi pengecekan missing value, penghapusan data duplikat, penanganan outlier, transformasi data kategorikal menggunakan Label Encoding, serta normalisasi data. Hasil proses normalisasi data ditunjukkan pada Gambar 5.1, sedangkan hasil transformasi data menggunakan Label Encoding ditunjukkan pada Gambar 5.2.

	Tahun	Bulan	Umur_Tanaman	Luas_Lahan	Hasil_Panen
0	0.0	0.0	1.0	0.0	0.437114
1	0.0	0.0	1.0	0.0	0.563874
2	0.0	0.0	1.0	0.0	0.409068
3	0.0	0.0	1.0	0.0	0.470472
4	0.0	0.0	1.0	0.0	0.710032

Gambar 5. 1 Normalisasi Data

	Kolom	Tipe Data
0	Tahun	int64
1	Bulan	int64
2	Hasil_Panen	int64
3	Umur_Tanaman	int64
4	Jenis_Tanah	int64
5	Urea	int64
6	NPK	int64
7	Ponska	int64
8	Borat	int64
9	Mutiara	int64
10	Musim	int64
11	Luas_Lahan	int64
12	Jenis_Tanah_Label	object
13	Musim_Label	object

Gambar 5. 2 Label Encoding fitur

Setelah tahap preprocessing selesai dilakukan, model Random Forest Regression dikembangkan dengan memanfaatkan data yang telah melalui tahap persiapan sebelumnya. Algoritma ini dipilih karena mampu menangani hubungan non-linear antarvariabel serta menghasilkan performa prediksi yang stabil pada berbagai jenis data.

Model yang telah dikembangkan selanjutnya diterapkan ke dalam aplikasi berbasis web dengan menggunakan Streamlit, sehingga dapat digunakan secara interaktif dan mudah diakses oleh pengguna. Implementasi sistem terdiri atas halaman beranda, halaman prediksi, halaman evaluasi, dan halaman visualisasi data.



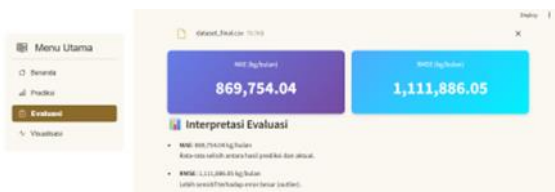
Gambar 5. 3 Halaman Beranda

Halaman beranda berfungsi sebagai tampilan awal aplikasi yang memberikan informasi mengenai tujuan dan fungsi sistem prediksi hasil panen kelapa sawit.



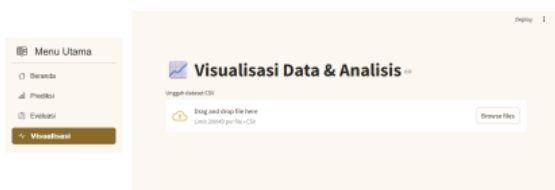
Gambar 5. 4 Halaman Prediksi

Halaman prediksi digunakan untuk memasukkan data berupa tahun panen, bulan panen, umur tanaman, jenis tanah, pemupukan, musim, dan luas lahan yang selanjutnya diproses oleh model untuk menghasilkan prediksi hasil panen.



Gambar 5. 5 Halaman Evaluasi

Halaman evaluasi menampilkan hasil pengukuran performa model menggunakan metrik evaluasi yang digunakan dalam penelitian.



Gambar 5. 6 Halaman Visualisasi

Halaman visualisasi digunakan untuk menampilkan grafik dan informasi pendukung yang membantu pengguna memahami pola data hasil panen secara lebih mudah.

Berdasarkan hasil pengujian, model Random Forest Regression terbukti mampu memperkirakan hasil panen dengan tingkat kesalahan yang tergolong rendah. Kinerja model dievaluasi menggunakan Mean Absolute Error (MAE), yakni sebuah metrik yang digunakan untuk menghitung rata-rata selisih mutlak antara nilai aktual dan nilai yang dihasilkan oleh model prediksi. Berdasarkan hasil pengujian diperoleh nilai MAE sebesar 869,75 kg/bulan.

Selain MAE, penelitian ini juga menggunakan Root Mean Squared Error (RMSE) digunakan untuk mengukur besarnya kesalahan yang dihasilkan oleh model prediksi. RMSE memberikan penalti yang lebih besar terhadap kesalahan prediksi yang bernilai tinggi, sehingga metrik ini mampu menggambarkan tingkat kestabilan model secara lebih komprehensif. Hasil pengujian menunjukkan nilai RMSE sebesar 1111,89 kg/bulan.

Nilai MAE dan RMSE yang diperoleh menunjukkan bahwa model mampu merepresentasikan pola historis hasil panen kelapa sawit dengan cukup baik. Dengan demikian, sistem yang dikembangkan dapat digunakan sebagai alat bantu dalam melakukan estimasi hasil panen serta mendukung proses pengambilan keputusan pada Kelompok Tani "ASRI".

SIMPULAN

Berdasarkan hasil penelitian yang telah dilaksanakan, dapat ditarik kesimpulan bahwa algoritma Random Forest Regression berhasil diterapkan dalam aplikasi prediksi hasil panen kelapa sawit berbasis web yang

dibangun menggunakan framework Streamlit. Proses preprocessing data yang meliputi pembersihan data, transformasi menggunakan Label Encoding, dan normalisasi data mampu menghasilkan data yang lebih siap untuk proses pemodelan. Berdasarkan hasil pengujian, model menunjukkan kinerja yang cukup memuaskan dengan perolehan nilai Mean Absolute Error (MAE) sebesar 869,75 kg/bulan dan Root Mean Squared Error (RMSE) sebesar 1111,89 kg/bulan. Nilai tersebut mengindikasikan bahwa model mampu mengenali dan mempelajari pola historis hasil panen kelapa sawit dengan tingkat deviasi kesalahan yang tergolong rendah. Selain itu, penerapan model ke dalam aplikasi berbasis Streamlit memungkinkan pengguna melakukan prediksi hasil panen secara interaktif dan dapat dimanfaatkan sebagai pendukung pengambilan keputusan pada Kelompok Tani "ASRI".

Bagi penelitian yang akan datang, disarankan untuk menggunakan volume data yang lebih besar dengan rentang periode pengamatan yang lebih panjang, sehingga model dapat mempelajari pola hasil panen secara lebih mendalam dan menghasilkan prediksi yang lebih optimal. Selain itu, penambahan variabel yang berpotensi memengaruhi hasil panen, seperti tingkat curah hujan, kondisi suhu udara, kelembapan atmosfer, serta intensitas paparan sinar matahari, dapat dilakukan untuk meningkatkan akurasi prediksi. Penelitian selanjutnya juga dapat membandingkan metode Random Forest Regression dengan algoritma lainnya, seperti XGBoost, Gradient Boosting, dan Support Vector Regression, guna memperoleh model dengan kinerja prediksi yang lebih baik. Pengembangan sistem juga dapat dilakukan dengan menambahkan fitur penyimpanan riwayat prediksi serta integrasi dengan data cuaca secara real-time guna meningkatkan kemampuan sistem dalam

mendukung perencanaan panen yang lebih efektif.

DAFTAR PUSTAKA

- K. Kafrawi, N. Hesti, S. Syatrawati, I. Rahim, and Z. Kumalawati, "Tingkat pertumbuhan dan produksi kelapa sawit (*Elaeis guineensis* Jacq.) pada berbagai topografi lahan," *Jurnal Galung Tropika*, vol. 12, no. 2, pp. 203–212, Sep. 2023. <https://doi.org/10.31850/jgt.v12i2.1109>
- A. Asnawi and R. Kurniawan, "Sistem informasi prediksi hasil panen kelapa sawit dengan menggunakan metode double exponential smoothing," *Jurnal Informatika Teknologi dan Sains*, vol. 7, no. 1, pp. 282–288, 2025
- F. Dhawi, A. Ghafoor, N. Almousa, S. Ali, and S. Alqanbar, "Predictive modelling employing machine learning, convolutional neural networks (CNNs), and smartphone RGB images for non-destructive biomass estimation of pearl millet (*Pennisetum glaucum*)," *Frontiers in Plant Science*, vol. 16, 2025. <https://doi.org/10.3389/fpls.2025.1594728>
- A. Oikonomidis, C. Catal, and A. Kassahun, "Deep learning for crop yield prediction: a systematic literature review," 2023. Taylor & Francis. <https://doi.org/10.1080/01140671.2022.2032213>
- Nurdiyanto Yusuf, "Prediksi produksi daging sapi di Indonesia menggunakan Random Forest Regression: analisis data 2018–2025"
- Farhanuddin, S. E. K. Sihombing, and Yahfizham, "Komparasi multiple linear regression dan random forest regression dalam memprediksi anggaran biaya manajemen proyek sistem informasi," *Journal of Computers and Digital Business*, vol. 3, no. 2, pp. 86–97, May 2024. <https://doi.org/10.56427/jcbd.v3i2.408>
- J. Hutahaeen and D. Yusup, "Perbandingan metode linear regression, random forest & K-nearest neighbor untuk prediksi produksi hasil panen padi di Provinsi Jawa Barat," 2024
- "Tugas akhir prediksi hasil panen kelapa sawit dengan menerapkan algoritma regresi linear berganda di PT Dendy Marker Indah Lestari Musi Rawas Utara"

“Perbandingan prediksi jumlah transaksi ojek online menggunakan regresi linier dan random forest”

E. Fitri and D. Riana, “Analisa perbandingan model prediction dalam prediksi harga saham menggunakan metode linear regression, random forest regression dan multilayer perceptron,” *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi*, vol. 6, no. 1, 2022. <https://doi.org/10.46880/jmika.Vol6No1.pp69-78>

E. M. Morse-McNabb, M. F. Hasan, and S. Karunaratne, “A multi-variable Sentinel-2 random forest machine learning model approach to predicting perennial ryegrass biomass in commercial dairy farms in Southeast Australia,” *Remote Sensing*, vol. 15, no. 11, 2023. <https://doi.org/10.3390/rs15112915>

W. Zhang and Y. Zhou, “Research on differentiated teaching model of Python programming based on learning data analysis,” 2023

J. M. Nápoles-Duarte, A. Biswas, M. I. Parker, J. P. Palomares-Baez, M. A. Chávez-Rojó, and L. M. Rodríguez-Valdez, “Stmol: A component for building interactive molecular visualizations within Streamlit web applications,” *Frontiers in Molecular Biosciences*, vol. 9, 2022. <https://doi.org/10.3389/fmolb.2022.990846>

A. T. Widiyatmoko, S. Butsianto, and A. Nugroho, “Penerapan machine learning untuk prediksi kenaikan harga beras premium menggunakan algoritma regresi linier,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 3, pp. 1125–1132, 2025. <https://doi.org/10.57152/malcom.v5i3.2123>

S. Handayani, B. I. Nugroho, and E. U. S. Sedyta Utami, “Perbandingan metode regresi linier dan exponential smoothing dalam memprediksi harga minyak goreng kemasan di Indonesia,” *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 3, pp. 1530–1538, 2025. <https://doi.org/10.31004/riggs.v4i3.2171>