

Optimasi Jadwal Tanam Padi di Kabupaten Tuban melalui Prediksi Curah Hujan Menggunakan Random Forest

Naili Nafa Khatirokimmah*, Mula Agung Barata, Sahri

Fakultas Sains dan Teknologi, Program Studi Teknik Informatika, Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia
Email: ¹nailinafaaa@gmail.com, ²mula.ab26@gmail.com, ³sahriunugiri@gmail.com
Email Penulis Korespondensi: nailinafaaa@gmail.com
Submitted 01-03-2026; Accepted 16-04-2026; Published 30-06-2026

Abstrak

Ketidakpastian cuaca akibat perubahan iklim semakin mengancam keberhasilan panen padi, terutama ketika petani masih mengandalkan ramalan cuaca tradisional yang tidak selalu akurat. Studi ini mengembangkan sistem pendukung keputusan untuk menentukan waktu penanaman padi berdasarkan prediksi curah hujan harian menggunakan Random Forest Regression. Data iklim harian dari BMKG di Tuban, Jawa Timur, untuk periode 2022–2025 digunakan sebagai dasar pelatihan, dengan penambahan fitur waktu seperti bulan, hari dalam tahun, dan musim untuk menangkap pola musiman. Di Jawa Timur, musim hujan biasanya berlangsung dari Oktober hingga April dan musim kemarau dari Mei hingga September, namun perubahan iklim telah menyebabkan pergeseran dalam waktu, durasi, dan intensitas curah hujan, sehingga klasifikasi musiman tradisional menjadi kurang dapat diandalkan untuk menentukan waktu penanaman optimal. Model ini diuji pada data 2025 dan menunjukkan kinerja yang lebih baik dibandingkan dengan model dasar. Model yang disesuaikan menghasilkan MAE sebesar 5,78 mm, RMSE sebesar 9,75 mm, dan koefisien determinasi (R^2) sebesar 0,177, yang lebih baik dibandingkan model dasar yang memiliki MAE sebesar 6,02 mm, RMSE sebesar 10,16 mm, dan R^2 sebesar 0,107. Meskipun nilai R^2 masih relatif rendah, penurunan MAE dan RMSE menunjukkan bahwa model yang disesuaikan lebih akurat dalam memprediksi curah hujan harian, terutama dalam rentang ringan hingga sedang, yang paling relevan untuk keputusan penanaman.

Kata Kunci: Pertanian Berbasis Data; Prediksi Curah Hujan; Random Forest Regression; SPK; Waktu Tanam Padi.

Abstract

Weather uncertainty due to climate change increasingly threatens rice harvest success, especially when farmers still rely on traditional forecasts that are not always accurate. This study developed a decision support system to determine the timing of rice planting based on daily rainfall predictions using Random Forest Regression. Daily climate data from the BMKG in Tuban, East Java, for the period 2022–2025 was used as the basis for training, with the addition of time features such as month, day of the year, and season to capture seasonal patterns. In East Java, the rainy season usually lasts from October to April and the dry season from May to September, but climate change has caused shifts in the timing, duration, and intensity of rainfall, making traditional seasonal classifications less reliable for determining the optimal planting time. The model was tested on 2025 data and showed improved performance compared to the baseline model. The tuned model produced an MAE of 5.78 mm, an RMSE of 9.75 mm, and a coefficient of determination (R^2) of 0.177, an improvement over the baseline, which had an MAE of 6.02 mm, an RMSE of 10.16 mm, and an R^2 of 0.107. Although the R^2 value is still relatively low, the decrease in MAE and RMSE indicates that the tuned model is more accurate in predicting daily rainfall, especially in the light to moderate range, which is most relevant for planting decisions.

Keywords: Data-Driven Agriculture; Rainfall Prediction; Random Forest Regression; DSS; Rice Planting Schedule.

1. PENDAHULUAN

Pentingnya Provinsi Jawa Timur dalam konteks pertanian padi nasional tidak dapat diabaikan. Data menunjukkan bahwa Jawa Timur secara konsisten masuk dalam daftar provinsi dengan kontribusi panen terbesar di Indonesia, menjadikannya pilar penting bagi ketahanan pangan dan stabilitas produksi padi nasional secara keseluruhan [1]. Dalam beberapa tahun terakhir, cuaca di Indonesia semakin tidak dapat diprediksi. Musim hujan sering datang terlambat, hujan deras dapat tiba-tiba turun di luar musim, sementara di tengah musim hujan yang seharusnya basah, beberapa daerah dilanda kekeringan. Situasi ini diperparah oleh fenomena *El Niño* yang berkepanjangan dari tahun 2023 hingga awal 2024, yang berdampak langsung pada produksi padi nasional, dengan banyak sawah gagal menghasilkan panen, dan Indonesia terpaksa meningkatkan impor beras [2],[3]. Hubungan antara ketidakpastian curah hujan dan produktivitas padi tidak selalu linier dan bervariasi antar wilayah, seperti yang ditunjukkan dalam studi di Kalimantan Barat yang menemukan bahwa perubahan pola curah hujan secara signifikan mengurangi hasil panen padi [4]. Petani yang selama ini mengandalkan pengalaman turun-temurun untuk menentukan waktu tanam kini menjadi bingung. Kalender tanam tradisional, yang dulu dapat diandalkan, kini tidak lagi sesuai dengan kondisi lapangan, karena pola cuaca yang dulu stabil kini menjadi tidak dapat diprediksi dan penuh ketidakpastian.

Dalam situasi seperti ini, informasi cuaca yang akurat, tepat waktu, dan dapat ditindaklanjuti bukan lagi kemewahan, melainkan kebutuhan mendesak. Ini bukan hanya soal menghindari banjir atau kekeringan, tetapi juga memastikan setiap musim tanam tetap produktif. Akses terhadap informasi cuaca yang tepat waktu dan relevan akan sangat membantu petani dalam merencanakan strategi budidaya mereka, sehingga meminimalkan risiko iklim [5]. Oleh karena itu, sistem agrometeorologi yang efektif perlu diterapkan untuk meningkatkan ketahanan pangan dan mendukung petani dalam menghadapi tantangan perubahan iklim [6]. Dan di sinilah waktu penanaman menjadi kunci: menanam terlalu awal berisiko kematian benih akibat kekeringan, sementara menanam terlalu lambat dapat menyebabkan tanaman terendam air atau diserang hama saat kondisi terlalu lembap. Oleh karena itu, pemahaman mendalam tentang pola cuaca

dan waktu tanam optimal sangat penting untuk memaksimalkan hasil pertanian dan menjaga ketahanan pangan di Indonesia [7],[8]. Tantangan iklim ini bukan lagi masalah teoretis. Studi kasus di berbagai wilayah Indonesia telah membuktikan bahwa anomali cuaca, seperti pergeseran awal musim hujan dan distribusi curah hujan yang tidak merata, memiliki dampak negatif langsung pada produktivitas padi. Kondisi ini menempatkan petani pada risiko tinggi, di mana keputusan waktu penanaman menjadi faktor penentu dalam menghindari kerugian [9]. Defiyanti, Sari, dan Padilah (2024) menunjukkan bahwa di Jawa Barat, waktu penanaman terbaik dapat ditentukan dengan memantau curah hujan per periode dekade (10 hari). Mereka menggunakan model ARIMA untuk memprediksi curah hujan, lalu menyarankan petani untuk mulai menanam pada periode dekadal ketiga November ketika curah hujan diperkirakan mencapai minimal 50 mm per periode dekadal [10]. Rekomendasi ini sejalan dengan prakiraan BMKG dan membuktikan bahwa prakiraan cuaca yang andal memang dapat diubah menjadi panduan praktis bagi petani [11].

Namun, pendekatan yang didasarkan pada ramalan musiman atau sepuluh hari masih memiliki keterbatasan. Surmaini dan Syahbuddin (2016) menekankan bahwa kriteria untuk memulai musim tanam tidak boleh hanya mengandalkan rata-rata, tetapi juga mempertimbangkan distribusi curah hujan harian untuk memastikan ketersediaan air yang cukup selama fase pertumbuhan kritis [12]. Di sinilah metode berbasis data seperti *machine learning* menunjukkan keunggulannya. Dengan menganalisis data historis secara mendalam, termasuk suhu, kelembapan, dan pola musiman, model ini dapat menangkap kompleksitas cuaca yang sulit dijelaskan menggunakan metode statistik tradisional. Misalnya, Dwiyanti dan Prianto (2023) berhasil membangun sistem prediksi cuaca di Jakarta menggunakan algoritma *Random Forest* dan data publik, menunjukkan bahwa *machine learning* dapat menyediakan informasi cuaca yang relevan tidak hanya untuk kawasan perkotaan tetapi juga untuk sektor pertanian [13]. Karya mereka membuktikan bahwa *Random Forest* sangat cocok untuk menangani sifat acak dan non-linier data meteorologi, menjadikannya sangat relevan untuk dukungan pengambilan keputusan agrometeorologi.

Untuk mengurangi risiko-risiko ini, inovasi dalam bentuk sistem pendukung keputusan sangat relevan. Dengan memanfaatkan kemajuan teknologi *machine learning*, algoritma seperti *Random Forest* menawarkan solusi yang menjanjikan. Metode ini dipilih karena keandalan yang telah teruji dalam memproses data kompleks dan non-linier seperti data iklim [14]. Ketangguhan algoritma *Random Forest* dalam mengelola data yang rumit dan mempertahankan tingkat akurasi yang tinggi juga telah dibuktikan secara lintas domain, termasuk dalam studi klasifikasi data kesehatan yang kompleks oleh Nisa et al. (2025) [15]. Dalam studi oleh Dwiyanti dan Prianto (2023), model ini terbukti sangat andal dalam membedakan sberbagai kondisi cuaca seperti cerah, hujan ringan, atau hujan deras dengan tingkat keberhasilan yang konsisten [13]. Bahkan dalam konteks global, Jeong dkk. (2016) menunjukkan bahwa *Random Forest* dapat memprediksi hasil panen lintas wilayah dengan akurasi tinggi, berkat kemampuannya menangkap hubungan non-linear antara cuaca dan pertumbuhan tanaman [16]. Di Indonesia, pendekatan serupa telah diterapkan untuk memprediksi produksi padi menggunakan berbagai skala data, dan terbukti efektif dalam mendukung perencanaan pertanian [17], [18], [19].

Inilah yang membedakan pendekatan peneliti dari studi sebelumnya. Defiyanti dkk. (2024) menggunakan prediksi dekadal, yang praktis tetapi terlalu kasar untuk keputusan harian [10]. Bayangkan dalam periode dekade tunggal, bisa saja ada hari-hari kering di tengah hujan, atau hujan deras berturut-turut yang berisiko banjir hal-hal yang tidak terlihat saat hanya melihat rata-rata 10 hari. Dengan *Random Forest*, kami dapat memberikan prediksi harian dan kemudian mengonversinya langsung menjadi rekomendasi operasional. Pendekatan serupa juga digunakan oleh Pratiwi dkk. (2024), yang mengubah prediksi curah hujan menjadi jadwal penanaman berdasarkan ambang batas tertentu, meskipun mereka menggunakan model ELM [18]. Berdasarkan hasil penelitian sebelumnya, *Random Forest* dipandang sebagai metode yang relevan dan potensial untuk diterapkan dalam prediksi curah hujan [20],[21].

Berdasarkan hal ini, studi ini bertujuan untuk menerapkan Regresi *Random Forest* untuk memprediksi curah hujan harian, kemudian menggunakannya sebagai dasar sistem pendukung keputusan penanaman padi. Berbeda dengan pendekatan agregat seperti ARIMA, sistem ini memberikan rekomendasi harian dalam kategori yang mudah dipahami, yaitu “Cocok untuk Penanaman” dan “Tunda Penanaman” berdasarkan ambang batas yang relevan dengan kebutuhan tanaman padi. Dengan mempertimbangkan interaksi kompleks antara variabel cuaca harian, pendekatan ini mengonversi informasi cuaca menjadi rekomendasi operasional yang dapat digunakan dalam praktik pertanian, mewujudkan visi Jain dan Ramesh (2020) untuk mengubah informasi cuaca menjadi rekomendasi keputusan pertanian yang konkret [22]. Dengan demikian, sistem ini diharapkan dapat diterapkan secara efektif oleh petani di lapangan, khususnya di wilayah sentra produksi padi seperti Provinsi Jawa Timur.

Kabupaten Tuban memiliki topografi yang beragam, salah satunya adalah wilayah dataran rendah yang terdapat di Kecamatan Soko. Wilayah ini didominasi oleh lahan pertanian yang sangat rentan terhadap genangan air dan berpotensi tinggi mengalami banjir ketika intensitas curah hujan meningkat. Oleh karena itu, sistem pendukung keputusan berbasis prediksi curah hujan harian ini secara khusus dikalibrasi dan diprioritaskan untuk dataran rendah di Kecamatan Soko. Di wilayah yang rawan banjir ini, pemantauan dan prediksi curah hujan harian menjadi sangat krusial, mengingat kalender tanam tradisional sudah tidak lagi relevan untuk memitigasi risiko gagal panen akibat luapan air dan genangan yang dapat merusak tanaman padi.

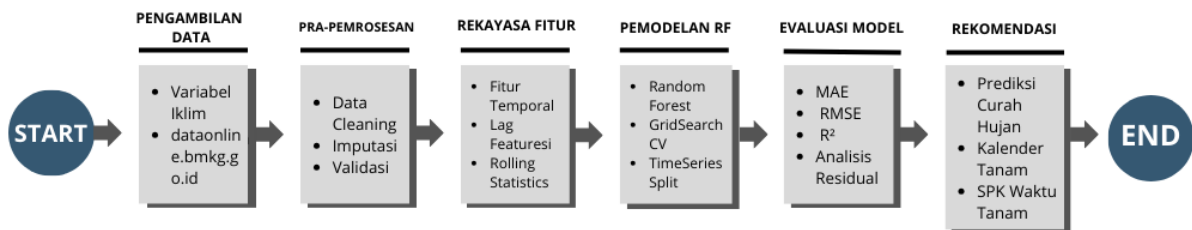
2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Metode penelitian ini disusun dalam rangkaian tahapan yang terstruktur dan saling berkesinambungan, dimulai dari pengelolaan data iklim hingga penjeremahan hasil prediksi menjadi panduan praktis bagi petani. Penelitian diawali dengan pengumpulan data iklim harian dari sumber resmi BMKG Stasiun Kabupaten di Tuban Jawa Timur untuk periode 2022–2025. Data tersebut kemudian menjalani proses pra-pemrosesan dan rekayasa fitur, termasuk pembersihan nilai tidak valid (seperti kode 8888 atau entri kosong), imputasi data hilang, serta pembuatan fitur baru seperti *lag*, *rolling average*, dan representasi musiman (bulan, hari dalam setahun, musim) untuk menangkap pola temporal yang relevan. Selanjutnya, data dibagi secara kronologis menjadi set latih dan uji, memastikan evaluasi model tetap realistis sesuai prinsip deret waktu tanpa kebocoran informasi masa depan ke dalam pelatihan. Pada tahap pemodelan, digunakan *Random Forest Regression* yang dilatih dan dioptimalkan melalui penyetulan hiperparameter (*hyperparameter tuning*) untuk mencapai keseimbangan antara akurasi dan kemampuan generalisasi.

Performa model dievaluasi menggunakan metrik statistik (MAE, RMSE, R^2) dan visualisasi residual guna memahami kekuatan serta keterbatasannya, terutama dalam menangkap pola curah hujan ekstrem. Hasil prediksi numerik kemudian dipetakan ke dalam kategori rekomendasi operasional berbasis intensitas curah hujan harian yaitu “Cocok Tanam” (jika ≤ 5 mm/hari) atau “Tunda Tanam” (jika > 5 mm/hari) yang relevan dengan kebutuhan budidaya padi. Terakhir, seluruh alur ini dirangkum dalam sebuah sistem pendukung keputusan berupa kalender tanam harian yang intuitif dan siap digunakan oleh petani untuk menentukan waktu tanam paling optimal berdasarkan kondisi cuaca prediktif.

Penelitian ini dirancang dengan alur kerja yang sistematis untuk memastikan validitas dan reliabilitas sistem yang dikembangkan. Secara keseluruhan, rangkaian proses penelitian digambarkan dalam Gambar 1 yang memvisualisasikan delapan tahapan utama mulai dari akuisisi data hingga diseminasi rekomendasi.



Gambar 1. Alur Kerja Penelitian

a. Pengumpulan dan Karakteristik Data

Penelitian ini menggunakan data iklim harian dari portal resmi BMKG untuk Stasiun Tuban, Jawa Timur, periode 8 Januari 2022–31 Desember 2025 (1.454 observasi). Dataset dibagi secara kronologis menjadi data latih (8 Januari 2022–16 Maret 2025) dan data uji (17 Maret 2025–31 Desember 2025). Variabel utama meliputi temperatur (minimum, maksimum, rata-rata), kelembapan udara, curah hujan, dan lama penyinaran matahari, ditambah ekstraksi fitur temporal (bulan, hari dalam tahun) untuk menangkap pola musiman.

b. Pra-pemrosesan Data dan Rekayasa Fitur

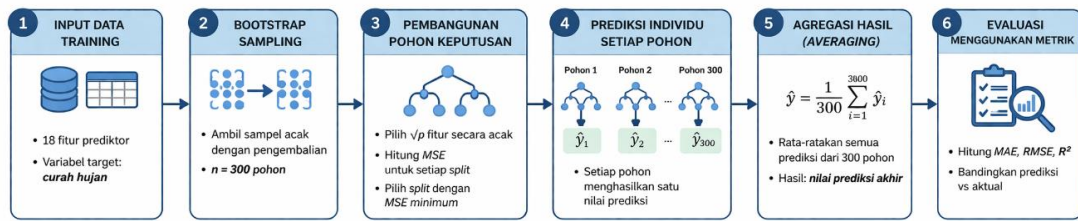
Data mentah dibersihkan dengan mengonversi nilai tidak valid (kode 8888, "-", kosong) menjadi *NaN*, lalu diimputasi: curah hujan menggunakan *forward/backward-fill*, variabel kontinu menggunakan median bergerak 7 hari. Validasi logis memastikan nilai berada dalam rentang fisik realistis. Rekayasa fitur menghasilkan 18 prediktor, meliputi fitur temporal, *lag* ($t-1$, $t-7$), dan *rolling mean* 7 hari, dengan curah hujan harian sebagai variabel target.

c. Pembagian Data Latih dan Uji Berdasarkan Waktu

Data dibagi secara kronologis untuk menjaga integritas deret waktu dan mencegah kebocoran informasi masa depan. Periode uji (Maret–Desember 2025) dipilih karena mencakup transisi musim lengkap (akhir hujan, puncak kemarau, awal hujan berikutnya), sehingga evaluasi model dapat menguji kinerja pada berbagai kondisi iklim yang representatif.

d. Pemodelan *Random Forest Regressor*

Model dikembangkan menggunakan *Random Forest Regressor* dengan prinsip *bagging* dan seleksi fitur acak per *node* untuk meminimalkan *Mean Squared Error* (MSE). Alur penerapan algoritma ini secara sistematis digambarkan pada Gambar 2.



Gambar 2. Penerapan algoritma *Random Forest* untuk prediksi curah hujan.

Diagram alir pada Gambar 2 mengilustrasikan tahapan penerapan algoritma: (1) *Input data training* dengan 18 fitur prediktor; (2) *Bootstrap sampling* untuk mengambil sampel acak; (3) Pembangunan pohon keputusan dengan pemilihan fitur acak dan kriteria MSE minimum; (4) Prediksi individual dari setiap pohon; (5) Agregasi hasil melalui *averaging*; dan (6) Evaluasi kinerja.

Hiperparameter dioptimalkan via *GridSearchCV* dengan *TimeSeriesSplit* ($n_splits=3$), mencakup max_depth [8,10,12], $min_samples_split$ [8,12,16], dan $min_samples_leaf$ [4,6,8]; $n_estimators$ ditetapkan 300. MAE dipilih sebagai metrik evaluasi utama karena lebih intuitif (satuan mm) dan *robust* terhadap *outlier*. Model terbaik dipilih berdasarkan MAE terendah pada validasi silang temporal, lalu diuji pada data yang belum pernah dilihat.

e. Evaluasi dan Perbandingan Model Kinerja

Kinerja model dievaluasi menggunakan tiga metrik regresi utama pada data uji. Misalkan n adalah jumlah observasi uji, y_t nilai aktual curah hujan (mm), \hat{y}_t nilai prediksi, dan \bar{y} rata-rata nilai aktual.

1. Root Mean Squared Error (RMSE) : Mengukur akar rata-rata kuadrat selisih aktual-prediksi; sensitif terhadap *outlier* karena memberikan penalti lebih besar pada kesalahan besar.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (A_t - F_t)^2} \quad (1)$$

2. Mean Absolute Error (MAE) : Mengukur rata-rata selisih absolut aktual-prediksi; lebih intuitif (satuan mm) dan *robust* terhadap nilai ekstrem, sehingga dipilih sebagai metrik utama.

$$MAE = \frac{1}{n} \sum_{t=1}^n |A_t - F_t| \quad (2)$$

3. Koefisien Determinasi (R^2) : Mengukur proporsi variansi data aktual yang dapat dijelaskan model. Nilai mendekati 1 menunjukkan kemampuan representasi pola yang baik.

$$R^2 = 1 - \frac{\sum_{t=1}^n (A_t - F_t)^2}{\sum_{t=1}^n (A_t - \bar{A})^2} \quad (3)$$

Interpretasi:

$R^2 = 1$ (sempurna),

$R^2 = 0$ (setara prediksi rata-rata),

$R^2 < 0$ (lebih buruk dari rata-rata, indikasi *overfitting*).

f. Pemetaan Hasil Prediksi Menjadi Rekomendasi Waktu Tanam

Hasil prediksi curah hujan harian (RR_t) dikonversi menjadi rekomendasi operasional dalam format kalender tanam yang intuitif. Rekomendasi harian diturunkan langsung dari intensitas curah hujan harian menggunakan ambang batas agronomis 5 mm/hari, yang merepresentasikan batas minimal ketersediaan air untuk perkecambahan benih padi tanpa risiko genangan berlebih. Kriteria pemetaan hasil prediksi menjadi rekomendasi waktu tanam disajikan pada Tabel 1 berikut.

Tabel 1. Kriteria pemetaan curah hujan harian menjadi rekomendasi operasional.

Kategori Rekomendasi	Ambang Batas Curah Hujan (RR_t)	Interpretasi Operasional
Cocok Tanam	$RR_t \leq 5$ mm/hari	Kondisi curah hujan ringan atau kering yang mendukung persiapan lahan, penanaman benih, dan proses perkecambahan tanpa risiko benih hanyut atau membusuk akibat genangan air.

Tunda Tanam	$RR_t > 5$ mm/hari	Curah hujan sedang hingga lebat yang berpotensi menyebabkan genangan, erosi tanah, serta menghambat pertumbuhan awal tanaman. Penanaman disarankan ditunda hingga kondisi cuaca lebih stabil.
-------------	--------------------	---

Berdasarkan kriteria pada Tabel 1, setiap nilai prediksi curah hujan harian dipetakan secara otomatis ke dalam salah satu kategori rekomendasi. Output akhir sistem berupa kalender tanam harian yang menampilkan rekomendasi biner sepanjang tahun, memungkinkan petani untuk merencanakan waktu tanam secara fleksibel dan adaptif terhadap dinamika cuaca prediktif di wilayah Tuban, Jawa Timur.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Penelitian ini menggunakan data iklim harian dari portal *online* resmi BMKG (<https://dataonline.bmkg.go.id>) stasiun Kabupaten Tuban, Jawa Timur. Data mencakup periode 8 Januari 2022 hingga 31 Desember 2025, menghasilkan total 1.454 observasi harian yang dibagi menjadi data latih (8 Januari 2022 – 16 Maret 2025) dan data uji (17 Maret 2025 – 31 Desember 2025). Variabel utama meliputi: Temperatur Minimum ($^{\circ}\text{C}$), Temperatur Maksimum ($^{\circ}\text{C}$), Temperatur Rata-Rata ($^{\circ}\text{C}$), Kelembapan Udara (%), Curah Hujan (mm), dan Lama Penyinaran Matahari (jam). Informasi temporal juga diekstraksi untuk mengidentifikasi pola musiman. Tabel 2 menunjukkan contoh data iklim harian untuk 5 hari pertama Januari 2022.

Tabel 2. Contoh data iklim harian BMKG Stasiun Tuban (1-5 Januari 2022).

Tanggal	Temperatur Minimum	Temperatur Maximum	Temperatur Rata-Rata	Kelembapan	Curah Hujan	Lama Penyinaran Matahari
01-01-2022	24,8	31,2	27,7	82	1,3	2,8
02-01-2022	24,8	31,2	27,4	87	4	3
03-01-2022	25	32,4	28,5	79	0	3,1
04-01-2022	23,6	32,2	27,2	82	0	7,8
05-01-2022	24,6	31,2	27,8	80	0	4,9

Data pada Tabel 2 menunjukkan variasi kondisi iklim harian di Tuban dengan temperatur 23,6-32,4 $^{\circ}\text{C}$, kelembapan 79-87%, curah hujan 0-4 mm, dan penyinaran matahari 2,8-7,8 jam/hari. Data primer dari sensor BMKG ini menjamin validitas dan relevansi untuk pengembangan sistem pendukung keputusan waktu tanam padi.

3.2 Pra-pemrosesan Data

Data iklim yang digunakan telah melalui serangkaian langkah pra-pemrosesan, termasuk pembersihan data untuk menangani nilai yang hilang atau anomali, normalisasi semua variabel (seperti suhu, kelembapan, dan durasi sinar matahari) ke skala seragam [0, 1], serta transformasi logaritmik curah hujan menggunakan fungsi \log_{1p} untuk menangani ketidakseimbangan distribusi akibat proporsi nilai nol yang tinggi (58%). Hasilnya adalah dataset yang konsisten, terstandarisasi, dan lebih seimbang secara statistik, siap digunakan untuk analisis lebih lanjut atau pemodelan prediktif. Statistik deskriptif dari dataset setelah proses pra-pemrosesan disajikan pada Tabel 3 berikut untuk memberikan gambaran komprehensif mengenai distribusi dan karakteristik masing-masing variabel.

Tabel 3. Statistik deskriptif hasil Pemrosesan data.

Variabel	mean	std	min	25%	50%	75%	max	proporsi nilai 0
Temperatur Minimum	0.56	0.15	0.00	0.47	0.57	0.65	1.00	0.00
Temperatur Maximum	0.56	0.13	0.00	0.50	0.57	0.64	1.00	0.00
Temperatur Rata2	0.53	0.15	0.00	0.42	0.52	0.62	1.00	0.00
Kelembapan	0.62	0.13	0.00	0.53	0.64	0.72	1.00	0.00
Curah Hujan	0.04	0.10	0.00	0.00	0.00	0.03	1.00	0.58
Lama Penyinaran Matahari	0.62	0.25	0.00	0.48	0.74	0.77	1.00	0.03
RR Log _{1p}	0.16	0.25	0.00	0.00	0.00	0.31	1.00	0.58

Tabel 3 menunjukkan statistik deskriptif dari seluruh variabel setelah melalui proses normalisasi dan transformasi. Seluruh variabel kontinu (temperatur, kelembapan, dan lama penyinaran matahari) telah dinormalisasi ke skala [0, 1], sehingga nilai *mean* berkisar antara 0.53 hingga 0.62 dengan standar deviasi (*std*) antara 0.13 hingga 0.25. Variabel curah hujan menunjukkan karakteristik yang berbeda dengan nilai *mean* yang sangat rendah (0.04) dan proporsi nilai nol sebesar 58%, yang mengonfirmasi sifat *zero-inflated* dan distribusi yang menceng kanan (*right-skewed*). Untuk mengatasi hal ini, transformasi \log_{1p} diterapkan pada variabel curah hujan (RR Log_{1p}), yang menghasilkan distribusi yang lebih seimbang dengan *mean* 0.16 dan standar deviasi 0.25. Nilai kuartil (25%, 50%, 75%) pada tabel memberikan informasi tentang sebaran data, di mana sebagian besar nilai curah hujan terkonsentrasi pada kisaran rendah (0.00-0.03), sementara nilai ekstrem (hujan lebat) mencapai 1.00 setelah normalisasi. Hasil pra-pemrosesan ini memastikan bahwa dataset siap

digunakan untuk pemodelan *Random Forest* dengan kualitas data yang terstandarisasi dan distribusi yang lebih representatif.

3.3 Evaluasi Kinerja Model: *Baseline vs Hyperparameter Tuning*

Model prediksi curah hujan harian awalnya dibuat dengan pengaturan standar (*baseline*), kemudian ditingkatkan melalui penyesuaian parameter (*hyperparameter tuning*) untuk mengurangi risiko *overfitting* dan meningkatkan kemampuan generalisasi. Proses ini penting karena model pohon keputusan seperti yang digunakan di sini sangat sensitif terhadap struktur data; tanpa batasan, model dapat menjadi terlalu kompleks dan hanya "menghafal" data pelatihan, daripada benar-benar mempelajari pola cuaca umum. Perbandingan kinerja model sebelum dan setelah penyesuaian parameter disajikan pada Tabel 4 berikut.

Tabel 4. Hasil Performa Model Prediksi Curah Hujan perbandingan *baseline* dan *tuned*.

Model	MAE (mm)	RMSE (mm)	R ²
Baseline	6.02	10.16	0.107
Tuned	5.78	9.75	0.177

Tabel 4 menunjukkan peningkatan kinerja model setelah proses *hyperparameter tuning* diterapkan. Kombinasi parameter terbaik yang diperoleh adalah: kedalaman pohon maksimum (*max depth*) 8, dengan minimal 8 titik data di setiap daun (*min samples leaf*) dan minimal 8 titik data untuk membagi cabang (*min samples split*). Pengaturan ini menjaga keseimbangan antara kompleksitas model dan kemampuannya untuk menangkap pola nyata tanpa terjebak pada *noise*. Berdasarkan Tabel 4, setelah penyesuaian, kinerja model menunjukkan perbaikan konsisten pada data uji:

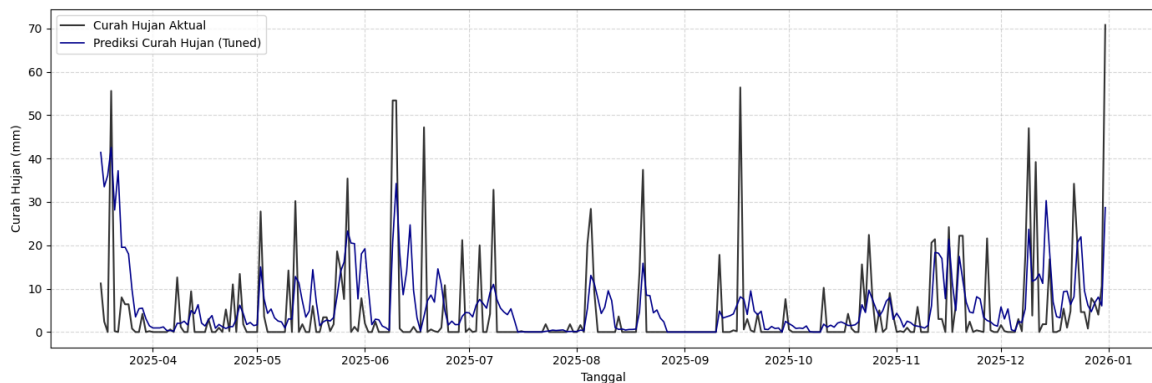
- MAE berkurang dari 6,02 mm menjadi 5,78 mm, artinya prediksi sekarang menyimpang rata-rata hanya sekitar 5,8 mm per hari dari nilai aktual.
- RMSE juga berkurang dari 10,16 mm menjadi 9,75 mm, menunjukkan bahwa model sedikit lebih baik dalam menangani prediksi ekstrem seperti hujan deras, karena RMSE memberikan penalti lebih besar terhadap kesalahan besar.
- Koefisien determinasi (R²) meningkat dari 0,107 menjadi 0,177, artinya model kini mampu menjelaskan 17,7% variasi curah hujan harian, lebih baik daripada versi *baseline* yang hanya menjelaskan 10,7%.

Meskipun nilai R² masih relatif rendah, hal ini wajar mengingat curah hujan harian sangat dipengaruhi oleh faktor atmosfer yang kompleks dan bersifat stokastik (acak). Dengan demikian, model *tuned* menjadi lebih stabil, sedikit lebih akurat, dan lebih dapat diandalkan untuk perkiraan operasional, terutama dalam konteks rekomendasi pertanian berbasis cuaca.

Selain itu, *Mean Absolute Percentage Error* (MAPE) sengaja tidak digunakan dalam evaluasi karena tidak cocok untuk data curah hujan harian. Curah hujan seringkali bernilai nol atau mendekati nol (misalnya 0,1 mm), dan ketika nilai aktual sangat kecil, kesalahan persentase dapat melonjak menjadi sangat besar atau bahkan tak terhingga jika nilai aktual benar-benar nol. Hal ini membuat MAPE menyesatkan dan tidak stabil sebagai ukuran kinerja. Sebagai gantinya, metrik seperti MAE dan RMSE lebih dapat diandalkan karena mengukur kesalahan dalam satuan asli (milimeter) tanpa terpengaruh oleh pembagian dengan angka yang hampir nol.

3.4 Visualisasi Hasil Prediksi: Deret Waktu Curah Hujan Aktual vs Prediksi

Untuk memberikan gambaran visual mengenai kemampuan model dalam menangkap pola curah hujan harian, Gambar 3 menyajikan perbandingan antara nilai aktual dan hasil prediksi model *Tuned Random Forest* sepanjang tahun 2025.



Gambar 3. Deret Waktu Curah Hujan Harian Tahun 2025 Aktual vs Prediksi

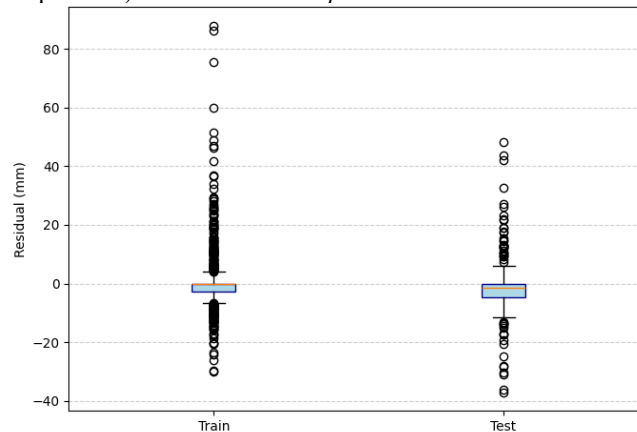
Gambar 3 memvisualisasikan perbandingan curah hujan aktual (garis hitam) dan prediksi model *Tuned Random Forest* (garis biru). Berdasarkan Gambar 3, model berhasil menangkap pola musiman dengan baik, terutama pada periode musim hujan (Oktober–Desember) di mana kedua garis menunjukkan tren kenaikan yang sejalan. Pada kondisi hujan ringan (<20 mm/hari), prediksi model juga cukup akurat mengikuti nilai aktual. Namun, model menunjukkan keterbatasan dalam memprediksi curah hujan ekstrem. Sebagai contoh, pada November 2025, nilai aktual mencapai >140 mm/hari,

sementara model hanya memprediksi 30–40 mm. Hal ini wajar karena algoritma *Random Forest* cenderung memprediksi nilai mendekati rata-rata kondisional dan kurang sensitif terhadap *outlier* ekstrem.

Meskipun demikian, untuk keperluan sistem pendukung keputusan pertanian, akurasi model pada kisaran curah hujan rendah hingga menengah sudah cukup memadai. Hal ini karena rekomendasi waktu tanam lebih bergantung pada kemampuan membedakan kondisi "kering/ringan" vs "basah/lebat" berdasarkan ambang batas 5 mm/hari, bukan pada ketepatan memprediksi nilai absolut curah hujan ekstrem.

3.5 Analisis Distribusi Kesalahan Berbasis Intensitas Hujan

Untuk memahami pola kesalahan prediksi model pada berbagai kondisi curah hujan, Gambar 4 menyajikan visualisasi distribusi residu (selisih aktual–prediksi) dalam bentuk *boxplot*.



Gambar 4. Boxplot (Error) per Kategori Intensitas Hujan

Gambar 4 menunjukkan distribusi residu untuk data latih dan data uji. Berdasarkan Gambar 4, sebagian besar kesalahan prediksi terkonsentrasi pada rentang -20 hingga $+20$ mm, mengindikasikan bahwa pada kondisi normal, model menyimpang dari nilai aktual sekitar ± 20 mm. Hal ini wajar mengingat fluktuasi tinggi pada data curah hujan harian.

Namun, terdapat beberapa *outlier* positif (di atas 40–80 mm) yang menunjukkan keterbatasan model dalam memprediksi curah hujan ekstrem (>100 mm/hari). Model cenderung *underestimate* pada kejadian hujan lebat, yang merupakan tantangan umum dalam peramalan cuaca berbasis *machine learning*.

Kesamaan pola *boxplot* antara data latih dan uji (median, IQR, dan rentang) mengindikasikan bahwa model tidak mengalami *overfitting* parah dan mampu mempertahankan konsistensi kinerja pada data baru. Meskipun demikian, adanya *outlier* ekstrem pada data uji mengingatkan bahwa kehati-hatian diperlukan jika prediksi ini digunakan untuk keputusan kritis seperti mitigasi banjir atau penjadwalan tanam. Dalam konteks pertanian, kesalahan prediksi besar dapat berdampak signifikan terhadap keberhasilan panen, sehingga rekomendasi sistem sebaiknya dipadukan dengan pemantauan cuaca lokal secara berkala.

3.6 Rekomendasi Waktu Tanam Padi Kota Tuban Tahun 2026

Sebagai output akhir sistem pendukung keputusan, Tabel 5 menyajikan contoh rekomendasi waktu tanam padi berbasis prediksi curah hujan harian untuk tahun 2026. Rekomendasi ini menggunakan ambang batas agronomis 5 mm/hari untuk menentukan kategori "Cocok Tanam" atau "Tunda Tanam".

Tabel 5. Rekomendasi Waktu Tanam Padi Tahun 2026 (sampel 3 hari pertama setiap bulan).

Bulan	Tanggal	Curah Hujan (mm)	Rekomendasi
Januari	1	8.35	Tunda Tanam
Januari	2	8.35	Tunda Tanam
Januari	3	8.35	Tunda Tanam
Februari	1	5.17	Tunda Tanam
Februari	2	5.17	Tunda Tanam
Februari	3	5.17	Tunda Tanam
Maret	1	5.18	Tunda Tanam
Maret	2	5.18	Tunda Tanam
Maret	3	5.18	Tunda Tanam
April	1	5.22	Tunda Tanam
April	2	5.17	Tunda Tanam
April	3	5.19	Tunda Tanam
Mei	1	4.97	Cocok Tanam
Mei	2	4.97	Cocok Tanam
Mei	3	4.97	Cocok Tanam

Juni	1	4.88	Cocok Tanam
Juni	2	4.88	Cocok Tanam
Juni	3	4.88	Cocok Tanam
Juli	1	4.93	Cocok Tanam
Juli	2	4.93	Cocok Tanam
Juli	3	4.93	Cocok Tanam
Agustus	1	4.92	Cocok Tanam
Agustus	2	4.92	Cocok Tanam
Agustus	3	4.92	Cocok Tanam
September	1	5.44	Tunda Tanam
September	2	5.44	Tunda Tanam
September	3	5.44	Tunda Tanam
Oktober	1	5.12	Tunda Tanam
Oktober	2	5.12	Tunda Tanam
Oktober	3	5.12	Tunda Tanam
November	1	5.03	Tunda Tanam
November	2	5.03	Tunda Tanam
November	3	5.03	Tunda Tanam
Desember	1	5.24	Tunda Tanam
Desember	2	5.24	Tunda Tanam
Desember	3	5.24	Tunda Tanam

Tabel 5 menunjukkan bahwa periode Mei hingga Agustus secara konsisten direkomendasikan sebagai "Cocok Tanam" dengan curah hujan prediksi stabil pada kisaran 4,88–4,97 mm/hari. Kondisi ini menyediakan kelembapan tanah yang cukup untuk perkecambahan tanpa risiko genangan berlebih. Sebaliknya, periode Januari–April dan September–Desember direkomendasikan sebagai "Tunda Tanam" karena curah hujan diprediksi >5 mm/hari (mencapai puncak 8,35 mm pada Januari), yang berisiko menyebabkan genangan, erosi, atau pembusukan benih.

Rekomendasi biner ini dirancang sebagai panduan strategis yang mudah diinterpretasikan petani untuk menyelaraskan jadwal tanam dengan pola cuaca prediktif. Dengan mengadopsi rekomendasi ini, petani dapat meningkatkan peluang keberhasilan panen dan meminimalkan kerugian akibat kondisi iklim yang tidak menguntungkan. Untuk implementasi operasional, sistem ini dapat dikembangkan menjadi antarmuka *mobile* atau *web-based* yang menampilkan kalender tanam interaktif dengan pembaruan prediksi berkala.

4. KESIMPULAN

Sistem pendukung keputusan berbasis *Random Forest Regression* berhasil memberikan rekomendasi yang andal dan didukung data untuk waktu penanaman padi. Model *Random Forest* telah melalui proses pembersihan data, pengolahan fitur lengkap (termasuk suhu, kelembapan, sinar matahari, serta fitur temporal dan lag), dan penyesuaian hiperparameter, menunjukkan kinerja stabil pada data uji 2025, dengan MAE sebesar 5,78 mm, RMSE sebesar 9,75 mm, dan R^2 sebesar 0,177. Meskipun nilai R^2 relatif rendah, mengingat sifat acak curah hujan harian, penurunan MAE dan RMSE dibandingkan dengan model dasar membuktikan bahwa prediksi lebih akurat, terutama pada rentang curah hujan ringan hingga sedang yang paling relevan bagi pertanian. Prediksi harian kemudian dikonversi menjadi rekomendasi operasional: "Cocok untuk Penanaman" jika curah hujan ≤ 5 mm/hari, dan "Tunda Penanaman" jika melebihi ambang batas tersebut. Hasil menunjukkan bahwa hanya periode Mei hingga Agustus yang secara konsisten memenuhi kriteria "Cocok untuk Penanaman", sementara bulan-bulan lain, termasuk Januari hingga April dan September hingga Desember, cenderung berisiko tinggi karena curah hujan melebihi 5 mm/hari. Sistem ini menawarkan panduan harian yang lebih presisi dibanding ramalan tradisional. Secara praktis, sistem ini diprioritaskan untuk dataran rendah Kecamatan Soko yang rentan banjir, di mana prediksi harian menjadi strategi adaptasi krusial untuk memitigasi gagal panen akibat genangan air, menggantikan kalender tanam tradisional yang tidak lagi relevan.

UCAPAN TERIMAKASIH

Kami ingin mengucapkan terima kasih kepada Program Studi Teknik Informatika dan Universitas Nahdlatul Ulama Sunan Giri Bojonegoro atas dukungan akademik, fasilitas, dan lingkungan ilmiah yang kondusif selama pelaksanaan penelitian ini. Ucapan terima kasih khusus juga disampaikan kepada para dosen dan staf yang telah memberikan bimbingan, masukan, dan dorongan sehingga penelitian ini dapat diselesaikan dan diterbitkan dalam jurnal ini.

REFERENCES

- [1] R. Rahim, A. Dela, R. Nurfalah, Y. Anggraeni, and S. Pasaribu, "Dinamika Ketahanan Pangan : Analisis Pengaruh Luas Panen Padi , Konsumsi Beras , Harga Beras , dan Jumlah Penduduk Terhadap Produksi Padi di Wilayah Sentra Padi di Indonesia Tahun 2017-2021," vol. 4, pp. 17083–17093, 2024.

- [2] R. A. Aldyan, "The impact of climate change on water resources and food security in Indonesia," *J. Law, Environ. Justice*, vol. 1, no. 1, pp. 50–63, 2023, doi: 10.62264/jlej.v1i1.2.
- [3] G. Manurung, I. Sastrodiharjo, and Y. Hendayana, "The Impact of Climate Change on Indonesia's Food Security Strategy Undertaken by Bulog in 2023," *Dinasti Int. J. Econ. Financ. Account.*, vol. 5, no. 3, pp. 1988–1994, 2024, doi: 10.38035/dijefa.v5i3.3204.
- [4] J. I. Lingkungan, F. Aditya, E. Gusmayanti, and J. Sudrajat, "Pengaruh Perubahan Curah Hujan terhadap Produktivitas Padi Sawah di Kalimantan Barat," vol. 19, no. 2, pp. 237–246, 2021, doi: 10.14710/jil.19.2.237-246.
- [5] K. Kim *et al.*, "in Agriculture," pp. 352–358.
- [6] G. Rusmayadi, U. Salawati, and D. O. Suparwata, "Analisis Sistem Agrometeorologi dalam Meningkatkan Ketahanan Pangan saat Kemarau," vol. 1, no. 03, pp. 143–150, 2023.
- [7] Sevina Yushinta Anjani, Bagus Setiawan, and Sofi Ayu Nur Martasari, "Dampak Perubahan Iklim Terhadap Ketahanan Pangan Di Indonesia," *J. Pendidik. Dan Ilmu Sos.*, vol. 2, no. 3, pp. 46–55, 2024, doi: 10.54066/jupendis.v2i3.1850.
- [8] J. M. Sihotang, Triman Tapi, and Nurhayati, "Dampak Perubahan Iklim terhadap Usaha Tani Padi: Studi Adaptasi Petani Kampung Prati Mulya, Distrik Prati, Kabupaten Manokwari," *J. Sustain. Agric. Ext.*, vol. 3, no. 2, pp. 107–119, 2025, doi: 10.47687/josae.v3i2.1646.
- [9] A. Bstract, "Analisis Perubahan Iklim Terhadap Produktivitas Padi Di Jawa Timur," vol. 13, no. 3, pp. 55–65, 2024.
- [10] S. Defiyanti *et al.*, "Optimasi Pertanian Padi : Peramalan Curah Hujan Berbasis Arima Rice Farming Optimization : Arima-Based Rainfall Forecasting For," vol. 11, no. 6, pp. 1377–1384, 2024, doi: 10.25126/jtiik.2024118682.
- [11] E. Surmaini, E. Susanti, Suciandini, M. R. Syahputra, and F. R. Fajary, "Use of the dry-spell seasonal forecast in crop management decisions," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 648, no. 1, 2021, doi: 10.1088/1755-1315/648/1/012092.
- [12] T. Padi and D. I. Indonesia, "Kriteria Awal Musim Tanam: Tinjauan Prediksi Waktu Tanam Padi Di Indonesia Onset Of Planting Season Criteria: Review Of Planting Time Prediction For Rice In Indonesia," no. Agus 2007, 2016, doi: 10.21082/jp3.v35n2.2016.p47-56.
- [13] Z. A. Dwiyanti and C. Prianto, "Prediksi Cuaca Kota Jakarta menggunakan Metode Random Forest : Studi Optimalitas," vol. 17, no. 2, pp. 127–137, 2023.
- [14] A. Primajaya *et al.*, "Random Forest Algorithm for Prediction of Precipitation," vol. 1, no. 1, pp. 27–31, 2018.
- [15] S. K. Nisa, M. A. Barata, and P. E. Yuwita, "Optimization of Random Forest Algorithm with SMOTE Method to Improve the Accuracy of Early Diabetes Prediction," vol. 12, no. 3, pp. 387–396, 2025, doi: 10.15294/sji.v12i3.22986.
- [16] J. P. Resop, N. D. Mueller, D. H. Fleisher, K. Yun, E. E. Butler, and D. J. Timlin, "Random Forests for Global and Regional Crop Yield Predictions Citation Random Forests for Global and Regional Crop Yield Predictions," 2016, doi: 10.1371/journal.pone.0156571.
- [17] T. Online, N. Nur, and F. Wajidi, "Jurnal Politeknik Caltex Riau Implementasi Algoritma Random Forest Regression untuk Memprediksi Hasil Panen Padi di Desa Minanga," vol. 9, no. 1, pp. 58–64, 2023.
- [18] I. Pratiwi, D. Bachtiar, A. Jauhari, M. Yusuf, F. A. Mufarroha, and D. Rosa, "Extreme Learning Machine Untuk Memprediksi Curah Hujan Dalam Penentuan Jadwal Tanam Padi," vol. 4, no. 1, pp. 34–42, 2024.
- [19] A. W. Wijayanto and S. R. Putri, "Estimating Rice Production using Machine Learning Models on Multitemporal Landsat-8 Satellite Images (Case Study: Ngawi Regency, East Java, Indonesia)," *Proc. - 2022 IEEE Int. Conf. Cybern. Comput. Intell. Cybern. 2022*, pp. 280–285, 2022, doi: 10.1109/CyberneticsCom55287.2022.9865364.
- [20] F. Collins, "Forecasting Excessive Rainfall with Random Forests and a Deterministic Convection-Allowing Model," no. M1, pp. 1693–1711, 2021, doi: 10.1175/WAF-D-21-0026.1.
- [21] A. Syahreza, N. K. Ningrum, and M. A. Syahrazy, "Perbandingan Kinerja Model Prediksi Cuaca: Random Forest, Support Vector Regression, dan XGBoost," *Edumatic J. Pendidik. Inform.*, vol. 8, no. 2, pp. 526–534, 2024, doi: 10.29408/edumatic.v8i2.27640.
- [22] S. Jain, D. Ramesh, and S. Member, "Machine Learning convergence for weather based crop selection," 2020, doi: 10.1109/SCEECS48394.2020.75.