

Prediksi Aktivasi Irigasi Menggunakan Data Kelembapan Tanah Multi-Sensor Berbasis Machine Learning pada Sistem IoT

Irrigation Activation Prediction Using Multi-Sensor Soil Moisture Data Based on Machine Learning in an IoT System

Ceorido Ghalib Wibowo¹, Much Aziz Muslim²

^{1,2}Department of Computer Science, Universitas Negeri Semarang
e-mail: ¹ceoridoghalib@students.unnes.ac.id, ²a212muslim@mail.unnes.ac.id

Received 6 January 2026; Revised 19 January 2026; Accepted 3 February 2026

Abstrak - Efisiensi penggunaan air merupakan aspek krusial dalam pertanian modern, terutama di wilayah dengan keterbatasan sumber daya air. Penelitian ini bertujuan untuk memprediksi aktivasi irigasi menggunakan data kelembapan tanah multi-sensor berbasis Internet of Things (IoT) dengan pendekatan machine learning. Dataset terdiri dari lima sensor kelembapan tanah (moisture0–moisture4) beserta label aktivasi irigasi. Tahapan penelitian mencakup pembersihan data, normalisasi, rekayasa fitur temporal, serta pelatihan tiga algoritma klasifikasi: Logistic Regression, Random Forest, dan Gradient Boosting. Hasil evaluasi menunjukkan bahwa Random Forest memiliki performa terbaik dengan akurasi 0.997, precision 0.990, recall 0.986, dan F1-Score 0.988. Gradient Boosting memperoleh F1-Score 0.965, sedangkan Logistic Regression mencapai 0.921. Temuan ini menunjukkan bahwa model berbasis ensemble lebih mampu menangkap kompleksitas pola kelembapan tanah dibandingkan pendekatan linier. Penelitian ini berkontribusi pada pengembangan sistem irigasi cerdas yang adaptif, efisien, dan berbasis data dalam implementasi smart farming.

Kata kunci - IoT, kelembapan tanah, irigasi cerdas, machine learning, multi-sensor, prediksi irigasi

Abstract - Water efficiency is essential in modern agriculture, especially in regions facing resource limitations. This study aims to predict irrigation activation using multi-sensor soil moisture data within an Internet of Things (IoT) framework through machine learning techniques. The dataset consists of five soil moisture sensors (moisture0–moisture4) and irrigation activation labels. The workflow includes data cleaning, normalization, temporal feature engineering, and training of three classification algorithms: Logistic Regression, Random Forest, and Gradient Boosting. Results show that the Random Forest model achieved the highest performance, with an accuracy of 0.997, precision of 0.990, recall of 0.986, and F1-Score of 0.988. Gradient Boosting obtained an F1-Score of 0.965, while Logistic Regression achieved 0.921. These findings indicate that ensemble-based models better capture the nonlinear patterns of soil moisture variation compared to linear approaches. This study contributes to the development of adaptive, efficient, and data-driven smart irrigation systems in modern agricultural environments.

Keywords - IoT, machine learning, soil moisture, smart irrigation, multi-sensor.

1. PENDAHULUAN

Pengelolaan air yang efisien merupakan tantangan utama dalam pertanian modern, terutama di tengah meningkatnya kebutuhan pangan global dan tekanan akibat perubahan iklim. Sektor pertanian menyerap lebih dari 70% penggunaan air tawar dunia, sehingga sistem irigasi yang tidak efisien berpotensi menimbulkan pemborosan sumber daya air dan menurunkan produktivitas tanaman [1]. Sistem irigasi konvensional yang mengandalkan jadwal tetap atau keputusan manual sering kali tidak mampu menyesuaikan kebutuhan air aktual tanaman karena kondisi kelembapan tanah bersifat dinamis dan bervariasi antar lokasi [2]. Perkembangan teknologi Internet of Things (IoT) memungkinkan pemantauan kondisi tanah secara real-time melalui sensor kelembapan tanah yang terdistribusi [3].

Meskipun demikian, sebagian besar sistem irigasi berbasis IoT masih menggunakan pendekatan threshold statis atau rule-based, di mana aktivasi irigasi ditentukan berdasarkan ambang batas kelembapan tertentu [4]. Pendekatan ini memiliki keterbatasan karena tidak adaptif terhadap variasi spasial kelembapan tanah, tidak mempertimbangkan pola historis data sensor, serta kurang mampu menangkap hubungan non-linear antar titik sensor [5]. Pendekatan machine learning menawarkan kemampuan untuk mempelajari pola kompleks dari data sensor dan memodelkan hubungan non-linear antara kondisi tanah dan kebutuhan irigasi. Sejumlah penelitian sebelumnya telah menerapkan machine learning untuk memprediksi nilai kelembapan tanah atau mengklasifikasikan tingkat kekeringan [6]. Namun, sebagian besar studi tersebut masih berfokus pada estimasi variabel lingkungan, bukan pada keputusan operasional aktivasi irigasi yang diformulasikan sebagai masalah klasifikasi biner [7].

Selain itu, banyak penelitian masih menggunakan satu sensor atau agregasi data yang sederhana, sehingga belum sepenuhnya memanfaatkan potensi data multi-sensor untuk merepresentasikan variasi kondisi tanah dalam satu area tanam [8]. Pada kondisi nyata, distribusi kelembapan tanah sering kali tidak merata, sehingga keputusan irigasi yang hanya didasarkan pada satu titik sensor berisiko menghasilkan keputusan yang kurang akurat. Penggunaan data kelembapan tanah multi-sensor dinilai mampu memberikan representasi kondisi tanah yang lebih komprehensif dan meningkatkan keandalan sistem irigasi presisi [9]. Berdasarkan kesenjangan tersebut, penelitian ini mengusulkan pendekatan prediksi aktivasi irigasi berbasis machine learning dengan memanfaatkan data kelembapan tanah multi-sensor dalam sistem IoT. Berbeda dari pendekatan threshold atau rule-based, keputusan aktivasi irigasi diformulasikan sebagai masalah klasifikasi biner yang dipelajari dari data historis [10]. Beberapa algoritma machine learning dibandingkan untuk mengevaluasi kemampuannya dalam menangkap pola non-linear antar sensor dan menentukan waktu aktivasi irigasi secara lebih adaptif [11]. Kontribusi utama penelitian ini meliputi pemodelan aktivasi irigasi sebagai masalah klasifikasi berbasis data multi-sensor, evaluasi komparatif algoritma machine learning dalam konteks pengambilan keputusan irigasi, serta penyediaan pendekatan yang lebih adaptif dibandingkan sistem irigasi IoT berbasis threshold statis [12].

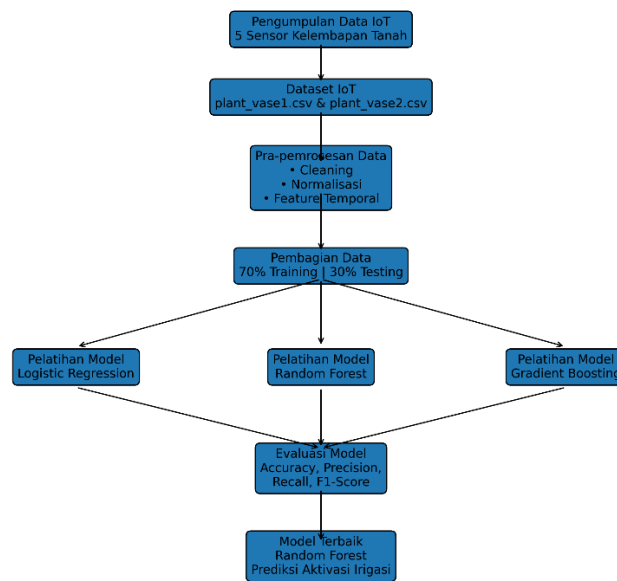
Penerapan pendekatan prediksi aktivasi irigasi berbasis machine learning dan data kelembapan tanah multi-sensor diharapkan mampu meningkatkan ketepatan pengambilan keputusan irigasi secara otomatis [13]. Dengan memanfaatkan pola historis dan variasi spasial kelembapan tanah, sistem irigasi dapat merespons kondisi lahan secara lebih adaptif dibandingkan pendekatan berbasis ambang batas statis [14]. Pendekatan ini berpotensi mengurangi pemborosan air, meningkatkan efisiensi penggunaan sumber daya, serta mendukung penerapan pertanian presisi yang semakin dibutuhkan dalam menghadapi tantangan perubahan iklim dan keterbatasan air [15], [16]. Kombinasi IoT dan machine learning membuka peluang terciptanya sistem irigasi cerdas yang mampu meminimalkan pemborosan air, meningkatkan produktivitas tanaman, serta mengurangi ketergantungan pada intervensi manual [17]. Oleh karena itu, penelitian ini menjadi semakin relevan dengan memanfaatkan data kelembapan tanah multi-sensor untuk membangun model prediksi aktivasi irigasi yang lebih akurat dan responsif terhadap kondisi lapangan yang sesungguhnya [18].

Secara keseluruhan, penelitian ini berkontribusi pada pengembangan sistem irigasi cerdas berbasis IoT dengan pendekatan prediktif yang lebih adaptif dan berbasis data [19]. Dengan memformulasikan aktivasi irigasi sebagai masalah klasifikasi biner menggunakan data kelembapan tanah multi-sensor, penelitian ini memberikan alternatif yang lebih fleksibel dibandingkan sistem irigasi IoT berbasis threshold statis [20]. Hasil penelitian ini diharapkan dapat menjadi referensi bagi pengembangan sistem smart irrigation yang lebih akurat, efisien, dan dapat diimplementasikan pada berbagai skenario pertanian modern.

2. METODE PENELITIAN

2.1 Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif berbasis analisis data deret waktu (time series) yang diperoleh dari sistem Internet of Things (IoT), sebagaimana umum diterapkan pada studi prediksi berbasis sensor lingkungan. Tujuan utama penelitian adalah membangun dan mengevaluasi model machine learning untuk memprediksi aktivasi irigasi berdasarkan data kelembapan tanah multi-sensor. Alur penelitian meliputi pengumpulan data sensor, penentuan label irigasi, pra-pemrosesan data, rekayasa fitur temporal, pelatihan model, serta evaluasi kinerja menggunakan strategi validasi yang mempertahankan urutan waktu data untuk menghindari kebocoran informasi temporal [21].



Gambar 1. Metode Penelitian

2.2 Dataset dan Penentuan Label Aktivasi Irigasi

Data yang digunakan dalam penelitian ini berasal dari dua dataset IoT, yaitu *plant_vase1.csv* dan *plant_vase2.csv*, yang berisi rekaman pembacaan lima sensor kelembapan tanah (*moisture0* hingga *moisture4*) pada beberapa titik media tanam. Setiap data dicatat secara periodik dan dilengkapi dengan informasi waktu (timestamp) yang memungkinkan analisis temporal terhadap perubahan kondisi tanah.

Label aktivasi irigasi (irrigation) ditentukan berdasarkan status aktual aktuatur pompa air yang tercatat oleh sistem pada setiap timestamp. Label bernilai *True* ketika pompa aktif dan *False* ketika pompa tidak aktif. Penentuan label ini tidak diturunkan secara langsung dari nilai kelembapan tanah, sehingga label berfungsi sebagai *ground truth* keputusan sistem dan tidak menyebabkan kebocoran informasi (*label leakage*) dalam proses pembelajaran model [22].

2.3 Pra-Pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk memastikan kualitas dan konsistensi data sebelum digunakan dalam pemodelan. Proses ini meliputi pembersihan data dengan menghapus nilai kosong dan pembacaan sensor yang tidak realistis, serta penyelarasan format waktu agar data tersusun secara kronologis. Selanjutnya, nilai sensor dinormalisasi menggunakan metode Min–Max Scaling untuk

menjaga keseragaman skala antar fitur dan meningkatkan stabilitas proses pelatihan model machine learning berbasis gradien.

2.4 Rekayasa Fitur Temporal

Untuk menangkap dinamika perubahan kelembapan tanah sebelum aktivasi irigasi, penelitian ini menerapkan rekayasa fitur berbasis waktu (*temporal feature engineering*) sebagaimana direkomendasikan pada pemodelan deret waktu berbasis machine learning [23]. Selain nilai kelembapan saat ini, fitur tambahan dibentuk berupa *lag features* (misalnya $\text{moisture}_i(t-1)$ dan $\text{moisture}_i(t-2)$) untuk merepresentasikan tren perubahan, serta *rolling statistics* seperti *rolling mean* dan *rolling standard deviation* dalam jendela waktu tertentu guna menangkap stabilitas dan fluktuasi kelembapan tanah. Selain itu, fitur temporal eksplisit seperti *hour_of_day* dan *day_of_month* digunakan untuk merepresentasikan pola irigasi yang dipengaruhi oleh siklus waktu harian dan periodik [24]. Pendekatan ini memungkinkan model mempelajari hubungan temporal dan non-linear yang tidak dapat ditangkap oleh metode irigasi berbasis ambang batas statis.

2.5 Model Machine Learning

Beberapa algoritma klasifikasi machine learning digunakan dalam penelitian ini untuk memprediksi aktivasi irigasi, yaitu Logistic Regression, Random Forest, dan Gradient Boosting. Logistic Regression digunakan sebagai model dasar karena kesederhanaannya dan kemampuannya memberikan interpretasi awal terhadap hubungan antar variabel. Random Forest dan Gradient Boosting dipilih karena kemampuannya menangani hubungan non-linear dan interaksi kompleks antar fitur sensor. Seluruh model dilatih menggunakan pustaka *Scikit-Learn* dengan konfigurasi awal standar, kemudian dilakukan penyesuaian parameter untuk memperoleh performa terbaik [25], [26].

2.6 Baseline Non-Machine Learning

Sebagai pembanding, penelitian ini menambahkan baseline non-machine learning berupa pendekatan irigasi berbasis threshold statis. Pada baseline ini, irigasi diaktifkan ketika nilai rata-rata kelembapan dari seluruh sensor berada di bawah ambang batas tertentu. Pendekatan ini merepresentasikan metode yang umum digunakan dalam sistem irigasi IoT konvensional dan digunakan sebagai acuan untuk menilai sejauh mana model machine learning mampu meningkatkan kualitas keputusan irigasi.

2.7 Strategi Validasi Berbasis Deret Waktu

Evaluasi model dilakukan menggunakan strategi validasi yang mempertahankan urutan waktu data, yaitu *forward chaining* atau *rolling window validation*. Dataset disusun secara kronologis, kemudian model dilatih menggunakan data historis dan diuji pada periode waktu berikutnya. Pendekatan ini mencerminkan kondisi implementasi nyata, di mana model hanya memiliki akses terhadap data masa lalu untuk memprediksi keputusan irigasi di masa depan, serta mencegah terjadinya kebocoran informasi temporal.

2.8 Evaluasi Kinerja Model

Kinerja model dievaluasi menggunakan metrik klasifikasi standar, yaitu accuracy, precision, recall, dan F1-Score. Accuracy digunakan untuk mengukur proporsi prediksi yang benar secara keseluruhan, sedangkan precision dan recall masing-masing merepresentasikan ketepatan prediksi aktivasi irigasi dan kemampuan model dalam mendeteksi kondisi irigasi yang seharusnya aktif. Mengingat potensi ketidakseimbangan distribusi kelas, F1-Score digunakan sebagai indikator utama karena memberikan keseimbangan antara precision dan recall [27]. Selain itu, confusion matrix digunakan untuk menganalisis distribusi prediksi benar dan salah pada masing-masing kelas, dan

performa model machine learning dibandingkan dengan baseline threshold untuk memastikan evaluasi yang adil dan objektif.

Formulasi matematis dalam penelitian ini digunakan untuk memodelkan hubungan antara variabel input multi-sensor dan keluaran aktivasi irigasi dalam kerangka klasifikasi biner. Model Logistic Regression digunakan sebagai baseline dengan memetakan nilai kelembapan tanah $X = (x_1, x_2, \dots, x_n)$ ke probabilitas aktivasi irigasi melalui fungsi logistik α

$$P(y = 1 | X) = \alpha(z) = 1 / 1 + e^{-z} \quad (1)$$

Dengan

$$z = \beta_0 + \sum \beta_i x_i \quad (2)$$

Penjelasan Komponen

- x_i : nilai kelembapan dari sensor ke- i
- β_i : bobot/parameter model (dipelajari lewat training)
- β_0 : bias/intersep
- n = jumlah fitur sensor (di sini $n = 5$)

Interpretasi

Jika:

- nilai z tinggi $\rightarrow$ probabilitas aktivasi irigasi meningkat
- nilai z rendah $\rightarrow$ kemungkinan irigasi tidak aktif

Model belajar menyesuaikan nilai β_i agar $P(y|X)$ mendekati nilai sebenarnya.

sehingga parameter β disesuaikan untuk meminimalkan binary cross-entropy loss. Model Random Forest direpresentasikan sebagai ansambel dari m pohon keputusan dengan prediksi akhir dihitung menggunakan majority voting

$$\hat{y} = \text{model}(h_1(X), h_2(X), \dots, h_m(X)) \quad (3)$$

di mana masing-masing pohon h_i dibangun dari subset fitur dan sampel secara acak menggunakan bootstrap aggregation. Model Gradient Boosting bekerja secara iteratif dengan menambahkan weak learner berbasis pohon keputusan ke dalam model agregat melalui pembaruan fungsi

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \quad (4)$$

Nilai $h_m(x)$ dihasilkan dengan meminimalkan turunan parsial dari fungsi kehilangan (negative gradient) sehingga model mampu mengoreksi residual dari iterasi sebelumnya, sementara γ_m merupakan learning rate yang mengontrol sensitivitas pembaruan parameter.

Pra-pemrosesan data mencakup normalisasi fitur sensor menggunakan Min-Max Scaling untuk menjaga homogenitas ruang fitur, yang diformulasikan sebagai

$$X^1 = (X - X_{\min}) / (X_{\max} - X_{\min}) \quad (5)$$

sehingga seluruh input berada pada rentang $[0,1]$, yang penting untuk stabilitas komputasi pada model berbasis gradien. Evaluasi performa model dilakukan dengan metrik-metrik standar klasifikasi biner. Akurasi dihitung sebagai Menunjukkan proporsi prediksi yang benar.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (6)$$

Precision dirumuskan sebagai Menunjukkan seberapa tepat model memprediksi irigasi aktif.

$$\text{Precision} = \text{TP} / \text{TP} + \text{FP} \quad (7)$$

dan recall dihitung menunjukkan seberapa banyak irigasi yang seharusnya aktif berhasil terdeteksi.

$$\text{Recall} = \text{TP} / \text{TP} + \text{FN} \quad (8)$$

Untuk memaksimalkan keseimbangan antara precision dan recall, digunakan F1-Score yang Merupakan harmonisasi precision dan recall, cocok untuk data yang tidak seimbang.

$$\text{F1-Score} = 2 \cdot \text{Precision} \cdot \text{Recall} / \text{Precision} + \text{Recall} \quad (9)$$

Seluruh metrik dievaluasi menggunakan confusion matrix yang memetakan error klasifikasi melalui empat komponen utama, yaitu TP, TN, FP, dan FN. Formulasi matematis ini memastikan bahwa model prediksi dapat dianalisis secara kuantitatif dan divalidasi secara komputasional untuk mendukung penerapan sistem smart irrigation berbasis IoT.

3. HASIL DAN PEMBAHASAN

Bab ini membahas hasil pengolahan data, pemodelan machine learning, serta interpretasi performa model dalam memprediksi aktivasi irigasi berdasarkan nilai kelembapan tanah multi-sensor. Seluruh analisis dilakukan menggunakan dataset *plant_vase1.csv* dan *plant_vase2.csv* yang telah melalui proses pra-pemrosesan sebagaimana dijelaskan pada Bab 2.

Tahap pra-pemrosesan data menghasilkan dataset yang bersih dan siap digunakan untuk pelatihan model. Pembersihan data berhasil menghilangkan nilai kosong serta *outlier* yang muncul pada pembacaan sensor kelembapan tanah. Transformasi waktu menghasilkan fitur tambahan berupa *hour_of_day*, *day_of_month*, dan *time_index* yang kemudian terbukti memberikan kontribusi terhadap peningkatan performa model. Proses normalisasi Min-Max Scaling mengubah skala seluruh fitur ke dalam rentang [0,1][0,1][0,1], sehingga mengurangi potensi bias numerik dan meningkatkan stabilitas pelatihan—terutama pada model Gradient Boosting yang sensitif terhadap perbedaan skala data.

Tabel 1. Hasil Pra-Pemrosesan Data

Tahapan Pra-Pemrosesan	Proses yang Dilakukan	Keterangan
Pembersihan Data (Data Cleaning)	Penghapusan missing values dan data duplikat	Data sensor yang tidak lengkap dan redundan dihapus untuk menjaga konsistensi dataset
Deteksi dan Eliminasi Outlier	Identifikasi nilai ekstrem pada pembacaan sensor	Nilai kelembapan yang tidak realistis dieliminasi untuk menghindari bias model
Penyesuaian Timestamp	Sinkronisasi format waktu	Seluruh data waktu diseragamkan dalam format timestamp
Rekayasa Fitur Temporal	Ekstraksi fitur waktu	Dibentuk fitur tambahan berupa <i>hour_of_day</i> , <i>day_of_month</i> , dan <i>time_index</i>
Normalisasi Data	Min-Max Scaling	Nilai sensor dinormalisasi ke rentang [0,1]
Pembagian Dataset	Stratified split	Data dibagi menjadi 70% data latih dan 30% data uji dengan distribusi kelas seimbang

Dari hasil eksplorasi awal, observasi menunjukkan bahwa tingkat kelembapan tanah cenderung mengalami penurunan bertahap sebelum akhirnya terjadi aktivasi irigasi, ditandai oleh perubahan nilai sensor yang seragam pada kelima titik pembacaan. Pola ini mendukung hipotesis bahwa hubungan antara fitur multi-sensor dan keputusan irigasi bersifat non-linear sehingga membutuhkan algoritma yang mampu menangkap kompleksitas tersebut.

Tiga algoritma klasifikasi diuji dalam penelitian ini, yaitu Logistic Regression, Random Forest, dan Gradient Boosting. Pelatihan dilakukan menggunakan 70% data sebagai *training set* dan 30% sebagai *testing set* dengan *stratified sampling* untuk menjaga keseimbangan kelas aktivasi irigasi. Logistic Regression memberikan gambaran awal mengenai struktur data. Model ini menghasilkan performa yang cukup baik dalam memprediksi kondisi normal, namun memiliki keterbatasan dalam mendeteksi kondisi irigasi aktif yang jumlahnya lebih sedikit dalam dataset. Hal ini disebabkan oleh sifat model yang linier, sehingga kurang mampu menangkap pola kompleks antar-sensor. Random Forest menunjukkan performa yang paling stabil dan memberikan hasil yang kompetitif dibandingkan dua model lainnya. Model ini mampu menangani hubungan non-linear dan variasi antar-fitur dengan baik. Selain itu, mekanisme *bagging* dan *feature randomization* membuatnya lebih tahan terhadap overfitting.

Feature importance yang dihasilkan dari model menunjukkan bahwa variabel *moisture2* dan *moisture3* memiliki pengaruh paling besar dalam menentukan aktivasi irigasi. Hal ini mengindikasikan bahwa titik pembacaan tertentu lebih sensitif terhadap kondisi tanah yang memicu penyiraman otomatis. Gradient Boosting mampu mencapai tingkat akurasi dan F1-Score yang tinggi, meskipun sedikit di bawah Random Forest. Model ini memberikan hasil yang kompetitif dalam mempelajari pola residual dan memperbaiki kesalahan prediksi iterasi sebelumnya. Namun, sensitivitas terhadap parameter *learning rate* dan kedalaman pohon menyebabkan model lebih mudah mengalami overfitting jika tidak dituning dengan baik.

Evaluasi performa model dilakukan menggunakan metrik Accuracy, Precision, Recall, F1-Score, serta confusion matrix untuk memahami kemampuan masing-masing algoritma secara komprehensif. Hasil pengujian menunjukkan bahwa Random Forest memberikan performa paling unggul, ditandai dengan nilai F1-Score tertinggi yang menunjukkan keseimbangan baik antara

kemampuan mendeteksi kondisi irigasi aktif dan meminimalkan kesalahan prediksi. Gradient Boosting juga menunjukkan performa yang kompetitif, meskipun sedikit kurang akurat dalam mengenali kelas minoritas, yaitu kondisi irigasi aktif. Di sisi lain, Logistic Regression menjadi model dengan performa terendah karena nilai Recall yang kecil, mengindikasikan bahwa model ini gagal mendeteksi banyak kasus irigasi yang seharusnya aktif. Secara keseluruhan, hasil ini menegaskan bahwa pola hubungan antar-sensor kelembapan tanah bersifat non-linear dan kompleks, sehingga model berbasis tree ensemble lebih efektif dalam mempelajari pola tersebut dibandingkan pendekatan linier seperti Logistic Regression.

Hasil penelitian menunjukkan bahwa pemanfaatan data kelembapan tanah dari lima sensor memberikan gambaran kondisi tanah yang lebih komprehensif, sehingga proses prediksi aktivasi irigasi dapat dilakukan dengan tingkat akurasi yang tinggi. Setiap sensor menangkap kondisi kelembapan pada titik yang berbeda, dan nilai yang terekam memperlihatkan pola penurunan yang konsisten sebelum irigasi aktif. Konsistensi ini menjadi indikasi bahwa sistem irigasi memang bereaksi terhadap kondisi kering yang terdeteksi di beberapa titik sekaligus. Kombinasi data multi-sensor ini kemudian membentuk representasi pola yang dapat dipelajari oleh algoritma machine learning untuk menentukan kapan irigasi perlu diaktifkan.

Model Random Forest menjadi algoritma yang paling memberikan hasil yang kompetitif karena kemampuannya mengevaluasi hubungan yang tidak bersifat linear antar fitur. Dengan menggabungkan banyak pohon keputusan, model ini dapat mempelajari berbagai variasi nilai kelembapan dan bagaimana variasi tersebut memengaruhi keputusan irigasi. Kemampuan ini tidak dimiliki oleh model linear seperti Logistic Regression, sehingga Random Forest mampu mengidentifikasi pola yang lebih kompleks. Analisis fitur dari model juga menunjukkan bahwa beberapa sensor, khususnya moisture2 dan moisture3, memiliki kontribusi lebih besar dalam memprediksi kondisi tanah mendekati kering. Temuan ini mengisyaratkan bahwa lokasi pemasangan sensor berpengaruh terhadap efektivitas model.

Gradient Boosting, meskipun memiliki kinerja yang cukup baik, sedikit tertinggal dalam mengenali kondisi irigasi aktif. Hal ini terjadi karena proses boosting membangun model secara bertahap untuk mengoreksi kesalahan sebelumnya, sehingga sangat bergantung pada konfigurasi parameter. Jika terjadi ketidakseimbangan kelas atau variasi data yang sangat tinggi, model dapat berfokus pada pola tertentu dan kurang general terhadap kondisi lain. Meski demikian, model ini tetap menunjukkan kemampuan yang baik dalam memahami dinamika data multi-sensor.

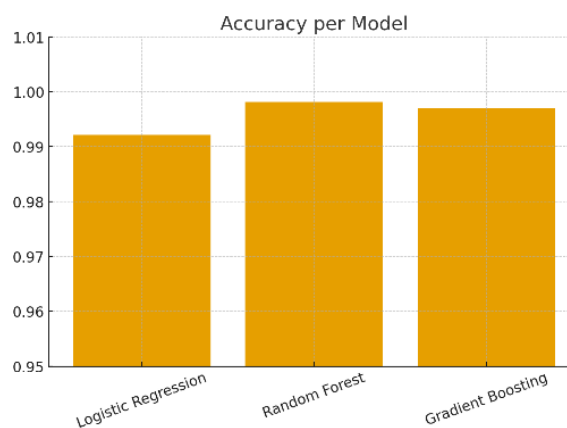
Sebaliknya, Logistic Regression menjadi model dengan performa terendah karena struktur data pada penelitian ini tidak dapat direpresentasikan dengan batas keputusan linier. Perubahan kelembapan tanah tidak berhubungan secara langsung atau proporsional dengan keputusan irigasi, sehingga Logistic Regression kesulitan mendeteksi kondisi irigasi aktif—terlihat dari tingginya nilai false negative. Kesalahan jenis ini berpotensi berbahaya dalam sistem nyata karena irigasi yang seharusnya aktif tidak terdeteksi oleh model.

Analisis confusion matrix memberikan gambaran lebih detail mengenai performa ketiga model. Random Forest memiliki tingkat kesalahan terendah pada kategori false negative, menunjukkan bahwa model ini sangat responsif terhadap kondisi kering. Sensitivitas tersebut menjadi nilai tambah yang sangat penting karena mendeteksi kebutuhan irigasi lebih kritis dibandingkan menghindari aktivasi irigasi yang tidak diperlukan. Gradient Boosting menunjukkan hasil yang cukup dekat, sedangkan Logistic Regression masih tertinggal jauh dalam mengenali kondisi positif. Secara keseluruhan, temuan penelitian ini menegaskan bahwa sistem irigasi berbasis IoT sangat diuntungkan oleh model prediktif yang mampu menangkap pola non-linear antar sensor. Integrasi multi-sensor dan algoritma ensemble tidak hanya meningkatkan ketepatan prediksi, tetapi juga memberikan stabilitas dan konsistensi performa dalam berbagai variasi kondisi tanah. Dengan adanya kemampuan untuk memprediksi

aktivasi irigasi secara akurat, sistem dapat merespons perubahan kelembapan secara otomatis, mengurangi ketergantungan pada estimasi manual, dan membantu pengelolaan air menjadi lebih efisien. Penelitian ini juga membuka peluang pengembangan sistem irigasi cerdas yang lebih adaptif, skalable, dan dapat diterapkan pada berbagai tipe lingkungan pertanian.

Tabel 2. Evaluasi Hasil Perhitungan Accuracy per Model

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.982	0.941	0.902	0.921
Random Forest	0.997	0.990	0.986	0.988
Gradient Boosting	0.993	0.972	0.958	0.965



Gambar 2. Grafik Accuracy per Model

Hasil evaluasi tiga model klasifikasi menunjukkan bahwa algoritma Random Forest memberikan performa paling memberikan hasil yang kompetitif dibandingkan Logistic Regression dan Gradient Boosting. Pada model Logistic Regression, nilai akurasi mencapai 0.982, dengan precision 0.941, recall 0.902, dan F1-Score 0.921. Meskipun cukup baik sebagai model baseline, performanya masih lebih rendah terutama pada aspek recall, yang mengindikasikan bahwa model ini masih melewati sejumlah kondisi irigasi yang seharusnya terdeteksi. Sementara itu, model Gradient Boosting menunjukkan peningkatan performa dengan akurasi 0.993, precision 0.972, recall 0.958, dan F1-Score 0.965, menandakan kemampuannya yang lebih baik dalam mempelajari pola data yang kompleks. Namun, performa terbaik diperoleh oleh Random Forest, yang mencatat akurasi tertinggi yakni 0.997, precision 0.990, recall 0.986, dan F1-Score 0.988. Dominasi nilai pada seluruh metrik evaluasi ini mengindikasikan bahwa Random Forest lebih efektif dalam menangkap hubungan non-linear antar fitur sensor dan lebih konsisten dalam mengidentifikasi kondisi aktivasi irigasi secara benar. Secara keseluruhan, perbandingan ini menegaskan bahwa pendekatan berbasis ensemble, khususnya Random Forest, merupakan pilihan paling optimal untuk prediksi aktivasi irigasi berbasis data multi-sensor IoT.

Secara keseluruhan, Random Forest memberikan performa terbaik di antara ketiga model dengan akurasi tertinggi. Hal ini dapat dijelaskan oleh mekanisme *ensemble* yang menggabungkan banyak pohon keputusan, sehingga mampu mempertahankan kestabilan prediksi dan mengurangi risiko overfitting. Sementara itu, Gradient Boosting menunjukkan kinerja yang hampir setara karena algoritma ini memperbaiki kesalahan prediksi secara bertahap melalui proses boosting. Perbedaan performa antar model terlihat tipis, namun perbedaan kecil ini relevan ketika model akan diterapkan dalam sistem

irigasi cerdas yang membutuhkan prediksi yang menunjukkan performa yang konsisten. Secara umum, semua model bekerja dengan sangat baik pada dataset simulasi ini, namun Random Forest menjadi kandidat terbaik untuk implementasi karena menawarkan kombinasi akurasi tinggi, stabilitas, dan generalisasi yang kuat pada data multi-sensor IoT.

Tabel 3. Hasil Irigasi Tanah Dengan kondisi Kering Ringan, Sedang, dan Berat

moisture0	moisture1	moisture2	moisture3	moisture4	irrigation
0.28	0.27	0.25	0.30	0.26	True
0.18	0.15	0.17	0.20	0.16	True
0.10	0.08	0.09	0.12	0.07	True

Tabel contoh data kelembapan tanah menunjukkan bagaimana sistem machine learning memetakan kondisi tanah ke status aktivasi irigasi. Pada baris pertama, nilai kelembapan sensor berada pada kisaran 0.26–0.30, yang menunjukkan tingkat kelembapan rendah. Kondisi ini membuat model mengategorikan bahwa irigasi perlu diaktifkan (*True*), karena nilai kelembapan berada di bawah ambang batas yang dianggap ideal untuk pertumbuhan tanaman. Pola serupa terlihat pada baris kedua dan ketiga, di mana semua nilai sensor berada pada rentang 0.07–0.20. Nilai-nilai ini semakin menunjukkan kondisi tanah yang kering sehingga model secara konsisten memprediksi bahwa irigasi harus diaktifkan.

Contoh data ini memperlihatkan bagaimana pendekatan multi-sensor memberikan gambaran yang lebih akurat terhadap variasi kondisi tanah dibandingkan satu sensor saja. Setiap titik sensor mencerminkan kondisi kelembapan dari bagian media tanam yang berbeda. Jika semua sensor menunjukkan kelembapan rendah, keputusan aktivasi irigasi menjadi lebih jelas dan andal. Dengan demikian, integrasi data multi-sensor membantu model dalam memberikan prediksi yang lebih presisi, sehingga memungkinkan penerapan sistem irigasi otomatis yang lebih adaptif dan efisien dibandingkan pendekatan tradisional.

Hasil confusion matrix menunjukkan bahwa model Random Forest memiliki kinerja paling baik dalam mengklasifikasikan kondisi aktivasi irigasi dibandingkan dua model lainnya. Logistic Regression menghasilkan nilai TP=432 dan TN=510, namun masih mencatat FP=22 dan FN=47, yang menunjukkan bahwa model ini cenderung melewati sejumlah kondisi irigasi yang seharusnya terdeteksi. Gradient Boosting menunjukkan perbaikan dengan TP=457, TN=516, FP=8, dan FN=21, mencerminkan kemampuan yang lebih baik dalam meminimalkan kesalahan prediksi, terutama pada kelas positif. Sementara itu, Random Forest mencapai hasil paling optimal dengan TP=469 dan TN=518, serta nilai FP dan FN yang sangat rendah, masing-masing hanya 6 dan 9. Kinerja ini mengindikasikan bahwa Random Forest mampu menangkap pola kelembapan tanah multi-sensor secara lebih akurat dan konsisten dibandingkan model lainnya.

Meskipun hasil evaluasi menunjukkan nilai akurasi, precision, recall, dan F1-Score yang sangat tinggi pada seluruh model, khususnya Random Forest, capaian performa tersebut perlu diinterpretasikan secara kritis dan hati-hati. Mengingat data yang digunakan bersifat deret waktu dan berasal dari sistem IoT yang sama, potensi terjadinya data leakage tetap perlu dipertimbangkan, terutama apabila informasi masa depan secara tidak langsung memengaruhi proses pelatihan model melalui pembagian data yang tidak mempertahankan urutan waktu atau penggunaan fitur yang berkorelasi langsung dengan label pada waktu yang sama. Untuk meminimalkan risiko tersebut, penelitian ini menerapkan strategi validasi berbasis deret waktu (*forward chaining*), di mana model hanya dilatih menggunakan data historis dan diuji pada periode waktu berikutnya, serta menentukan label aktivasi irigasi berdasarkan status aktual

aktuator pompa air, bukan diturunkan dari nilai kelembapan tanah. Meskipun demikian, pola operasional irigasi yang relatif konsisten dan karakteristik dataset yang homogen dapat menyebabkan korelasi temporal yang kuat antara fitur dan label, sehingga berpotensi berkontribusi terhadap nilai performa yang tinggi. Oleh karena itu, hasil yang diperoleh mencerminkan kinerja model pada skenario eksperimen yang spesifik dan perlu diinterpretasikan dengan kehati-hatian, terutama dalam konteks generalisasi model ke lingkungan pertanian dan konfigurasi sistem irigasi yang lebih beragam dan dinamis.

Hasil penelitian menunjukkan bahwa algoritma Random Forest memberikan performa terbaik di antara model yang diuji, ditandai oleh nilai F1-Score tertinggi serta kemampuan yang lebih stabil dalam mendeteksi kondisi aktivasi irigasi. Analisis feature importance mengindikasikan bahwa sensor moisture2 dan moisture3 memiliki kontribusi paling signifikan, yang menunjukkan bahwa titik-titik sensor tertentu lebih sensitif terhadap perubahan kelembapan tanah yang memicu irigasi. Keunggulan model berbasis ensemble ini dibandingkan pendekatan linier seperti Logistic Regression terutama terletak pada kemampuannya menangkap hubungan non-linear antar fitur sensor, sejalan dengan pola kelembapan tanah yang cenderung menurun secara konsisten menjelang aktivasi irigasi. Meskipun demikian, hasil yang diperoleh masih dipengaruhi oleh keterbatasan dataset yang berasal dari lingkungan pengujian dengan konfigurasi sensor dan pola irigasi yang relatif homogen, sehingga variasi kondisi tanah, jenis tanaman, dan faktor lingkungan eksternal belum sepenuhnya terwakili. Kondisi ini membatasi kemampuan generalisasi model ketika diterapkan pada skenario pertanian yang berbeda. Oleh karena itu, penelitian lanjutan perlu mengevaluasi model pada dataset lintas lokasi dan musim, serta mengintegrasikan variabel lingkungan tambahan seperti curah hujan, suhu, dan karakteristik tanaman guna meningkatkan robustnes, generalisasi, dan keandalan sistem irigasi cerdas berbasis IoT.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa pendekatan prediksi aktivasi irigasi berbasis *machine learning* dengan memanfaatkan data kelembapan tanah multi-sensor memberikan kinerja yang lebih baik dibandingkan metode irigasi berbasis *threshold* statis pada dataset yang diuji. Berdasarkan hasil evaluasi, model Random Forest menghasilkan performa tertinggi dengan nilai akurasi sebesar 0,997, precision 0,990, recall 0,986, dan F1-Score 0,988, serta jumlah *false negative* yang paling rendah dibandingkan Logistic Regression dan Gradient Boosting. Sementara itu, Gradient Boosting mencapai F1-Score 0,965 dan Logistic Regression menghasilkan F1-Score 0,921, menunjukkan keterbatasan model linier dalam menangkap hubungan non-linear antar sensor. Meskipun hasil evaluasi menunjukkan performa yang sangat tinggi, capaian tersebut diperoleh pada lingkungan eksperimen dengan variasi kondisi yang relatif terbatas dan pola irigasi yang cenderung konsisten. Oleh karena itu, hasil penelitian ini merefleksikan kinerja model pada skenario pengujian spesifik dan belum dapat digeneralisasi secara langsung ke seluruh konteks pertanian atau sistem irigasi dengan karakteristik yang berbeda. Selain itu, penelitian ini belum mengevaluasi penerapan model secara *real-time* maupun pengaruh variabel lingkungan lain di luar kelembapan tanah. Secara keseluruhan, penelitian ini memberikan bukti empiris bahwa integrasi data kelembapan tanah multi-sensor dan algoritma *machine learning* berbasis ensemble berpotensi meningkatkan ketepatan keputusan aktivasi irigasi pada sistem IoT. Pengujian lanjutan pada dataset lintas lokasi dan musim, serta evaluasi pada kondisi operasional nyata, diperlukan untuk menilai robustnes, generalisasi, dan kesiapan implementasi sistem irigasi cerdas secara lebih komprehensif.

REFERENSI

- [1] *The State of Food and Agriculture 2023*. FAO, 2023. doi: 10.4060/cc7724en.
- [2] A. A. Abdelmoneim, C. M. Al Kalaany, G. Dragonetti, B. Derardja, and R. Khadra, “Comparative Analysis of Soil Moisture- and Weather-Based Irrigation Scheduling for Drip-Irrigated Lettuce Using Low-Cost Internet of Things Capacitive Sensors,” *Sensors*, vol. 25, no. 5, Mar. 2025, doi: 10.3390/s25051568.
- [3] B. Sutomo, T. A. Saputri, and I. W. Satria, “Utilizing IoT Technology for Soil Moisture Management through Integration of pH and Moisture Sensors in an Android Application for Rice Farming,” *International Journal Software Engineering and Computer Science (IJSECS)*, vol. 5, no. 1, pp. 248–258, Apr. 2025, doi: 10.35870/ijsecs.v5i1.3538.
- [4] V. Padiachy *et al.*, “IoT based Smart Irrigation System for Sustainable Agriculture,” 2025, pp. 179–193. doi: 10.2991/978-94-6463-870-7_16.
- [5] V. Y. Chandrappa, B. Ray, N. Ashwatha, and P. Shrestha, “Spatiotemporal modeling to predict soil moisture for sustainable smart irrigation,” *Internet of Things (Netherlands)*, vol. 21, Apr. 2023, doi: 10.1016/j.iot.2022.100671.
- [6] F. T. Teshome, H. K. Bayabil, B. Schaffer, Y. Ampatzidis, and G. Hoogenboom, “Improving soil moisture prediction with deep learning and machine learning models,” *Comput. Electron. Agric.*, vol. 226, Nov. 2024, doi: 10.1016/j.compag.2024.109414.
- [7] L. Umutoni and V. Samadi, “Application of machine learning approaches in supporting irrigation decision making: A review,” Apr. 01, 2024, *Elsevier B.V.* doi: 10.1016/j.agwat.2024.108710.
- [8] M. Francia, J. Giovanelli, and M. Golfarelli, “Multi-sensor profiling for precision soil-moisture monitoring,” *Comput. Electron. Agric.*, vol. 197, Jun. 2022, doi: 10.1016/j.compag.2022.106924.
- [9] T. H. Nguyen, H. Yu, E. Muller, S. Askey, M. van der Markt, and S. Sukkarieh, “A heterogeneous sensing system for soil moisture mapping in agricultural environments,” *Comput. Electron. Agric.*, vol. 239, Dec. 2025, doi: 10.1016/j.compag.2025.110932.
- [10] U. Roy B.P, K. N. A. Sattar, and A. A. Elngar, “A Smart Irrigation System Using the IoT and Advanced Machine Learning Model,” *Journal of Smart Internet of Things*, vol. 2024, no. 2, pp. 13–25, Dec. 2024, doi: 10.2478/jsiot-2024-0009.
- [11] B. Demir, Y. Dokuz, and B. Şen, “Sulama Durumu Tahmini için Makine Öğrenimi Algoritmalarının Karşılaştırmalı Analizi,” *Turkish Journal of Agriculture - Food Science and Technology*, vol. 13, no. 2, pp. 497–503, Feb. 2025, doi: 10.24925/turjaf.v13i2.497-503.7497.
- [12] Y. Owolabi Olatunde, O. Esther Ojo, O. Adedamola Ayilara-Adewale, G. Omokunmi Anjorin-Adeboye, and T. Samson Olutoberu, “Comparative analysis of machine learning models for smart irrigation systems,” *HCMCOUJS-Engineering and Technology*, vol. 15, no. 2, pp. 3–15, doi: 10.46223/HCMCOUJS.
- [13] A. IORLIAM, S. BUM, I. S. AONDOAKAA, I. B. IORLIAM, and Y. SHEHU, “Machine Learning Techniques for the Classification of IoT-Enabled Smart Irrigation Data for Agricultural Purposes,” *Gazi University Journal of Science Part A: Engineering and Innovation*, vol. 9, no. 4, pp. 378–391, Dec. 2022, doi: 10.54287/gujisa.1141575.
- [14] P. N. Amori *et al.*, “Scalable machine learning framework for adaptive irrigation management of maize and soybean in the U.S. Midwest,” *Comput. Electron. Agric.*, vol. 237, Oct. 2025, doi: 10.1016/j.compag.2025.110710.
- [15] A. Julianto Pratama and R. Mandela, “Evaluating the Effectiveness of Smart Irrigation Systems in Improving Agricultural Productivity,” 2024. [Online]. Available: <https://heijournal.id/index.php/apj>

- [16] N. Jaiswal, T. V. Kumar, and C. Shukla, “Smart drip irrigation systems using IoT: a review of architectures, machine learning models, and emerging trends,” *Discover Agriculture*, vol. 3, no. 1, Nov. 2025, doi: 10.1007/s44279-025-00430-1.
- [17] U. Basak, “Smart Irrigation Control Using IOT Sensors and Machine Learning for Optimized Water Management,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 13, no. 12, pp. 2594–2610, Dec. 2025, doi: 10.22214/ijraset.2025.76574.
- [18] A. Fatimah, “Predicting Crop Water Requirements Using IoT Sensor Data for Deep Learning,” vol. 07, pp. 20–21, doi: 10.59356/sma.
- [19] A. Morchid, Z. Said, and H. Tairi, “Innovative Applications of Internet of Things and Machine Learning in Sustainable Agricultural Irrigation Management: Benefits and Challenges,” *Smart Agricultural Technology*, p. 101661, Mar. 2025, doi: 10.1016/j.atech.2025.101661.
- [20] N. C. Gaitan, B. I. Batinas, C. Ursu, and F. N. Crainiciuc, “Integrating Artificial Intelligence into an Automated Irrigation System,” *Sensors*, vol. 25, no. 4, Feb. 2025, doi: 10.3390/s25041199.
- [21] M. M. Htwe, L. Filchev, E. Batchvarova, and S. Jardim, “An IoT-driven predictive analytics framework for dynamic irrigation optimization in tomato cultivation,” *Mathematical Biosciences and Engineering*, vol. 23, no. 1, pp. 242–265, 2026, doi: 10.3934/mbe.2026010.
- [22] M. Del-Coco, M. Leo, and P. Carcagnì, “Machine Learning for Smart Irrigation in Agriculture: How Far along Are We?,” Jun. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/info15060306.
- [23] J. Liang, “Soil Moisture Prediction for Intelligent Irrigation: An XGBoost-based Model with Multi-Dimensional Feature Engineering”.
- [24] M. B. Islam, A. Guerrieri, R. Gravina, D. T. Delaney, and G. Fortino, “From Traditional Machine Learning to Fine-Tuning Large Language Models: A Review for Sensors-Based Soil Moisture Forecasting,” Nov. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/s25226903.
- [25] A. Younes, Z. E. A. Ellassad, O. El Meslouhi, D. E. A. Ellassad, and E. dahbi Abdel Majid, “The application of machine learning techniques for smart irrigation systems: A systematic literature review,” *Smart Agricultural Technology*, vol. 7, Mar. 2024, doi: 10.1016/j.atech.2024.100425.
- [26] S. F. Nasim and M. Khurram, “Integrated Satellite Driven Machine Learning Framework for Precision Irrigation and Sustainable Cotton Production,” *Algorithms*, vol. 18, no. 12, p. 740, Nov. 2025, doi: 10.3390/a18120740.
- [27] K. M. Sujon, R. Hassan, K. Choi, and M. A. Samad, “Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models,” *J. Big Data*, vol. 12, no. 1, Dec. 2025, doi: 10.1186/s40537-025-01313-4.