



## Memprediksi Jenis Tanaman Berdasarkan Parameter Lingkungan Menggunakan Algoritma Decision Tree dan Random Forest Berbasis Web

Nila Aulia<sup>1\*</sup>, Muhammad Rizky<sup>2</sup>, Chaerul Shaleh<sup>3</sup>

<sup>1-3</sup>Information Technology, Universitas Bina Sarana Informatika, Indonesia

\*Penulis korespondensi: [nilaaulia496@gmail.com](mailto:nilaaulia496@gmail.com)<sup>1</sup>

**Abstract.** *The advancement of information technology and artificial intelligence has created new opportunities to enhance efficiency and accuracy in the agricultural sector. One major challenge faced by farmers is determining the most suitable crop type based on environmental factors such as temperature, humidity, soil pH, rainfall, and light intensity. This study aims to develop a web-based crop prediction system using Decision Tree and Random Forest algorithms to support farmers in making data-driven decisions. The research methodology includes environmental data collection, data preprocessing, model training using both algorithms, performance evaluation based on accuracy metrics, and deployment of the best-performing model into a web application. The system allows users to input environmental parameters and obtain real-time crop predictions instantly. The novelty of this study lies in integrating both Decision Tree and Random Forest algorithms into a single interactive web platform, providing not only accurate predictions but also easy accessibility for users. Experimental results indicate that the Random Forest algorithm achieves higher accuracy than the Decision Tree in crop classification. Therefore, this system can serve as an effective tool for farmers and researchers to identify suitable crops under specific environmental conditions.*

**Keywords:** *Crop Prediction; Decision Tree; Machine Learning; Random Forest; Web-Based System*

**Abstrak.** Perkembangan teknologi informasi dan artificial intelligence membuka peluang baru untuk meningkatkan efisiensi dan akurasi di sektor pertanian. Salah satu permasalahan utama yang dihadapi petani adalah menentukan jenis tanaman yang paling sesuai berdasarkan faktor lingkungan seperti suhu, kelembapan, pH tanah, curah hujan, dan intensitas cahaya. Penelitian ini bertujuan mengembangkan sistem prediksi tanaman berbasis web menggunakan algoritma Decision Tree dan Random Forest untuk membantu petani membuat keputusan yang lebih berbasis data. Metodologi penelitian meliputi pengumpulan data lingkungan, preprocessing data, pelatihan model menggunakan kedua algoritma, evaluasi kinerja berdasarkan metrik akurasi, serta implementasi model terbaik ke dalam aplikasi web. Sistem ini dirancang agar pengguna dapat memasukkan parameter lingkungan dan memperoleh prediksi jenis tanaman secara cepat dan real-time. Kebaruan penelitian terletak pada integrasi algoritma Decision Tree dan Random Forest dalam satu platform web interaktif yang tidak hanya menyediakan prediksi yang akurat, tetapi juga mudah diakses oleh pengguna. Hasil eksperimen menunjukkan bahwa algoritma Random Forest memiliki akurasi lebih tinggi dibandingkan Decision Tree dalam klasifikasi tanaman. Dengan demikian, sistem ini dapat menjadi alat yang efektif bagi petani dan peneliti untuk menentukan tanaman yang sesuai dengan kondisi lingkungan tertentu.

**Kata kunci:** Hutan Acak; Pembelajaran Mesin; Pohon Keputusan; Prediksi Tanaman; Sistem Berbasis Web

### 1. LATAR BELAKANG

Di era pertanian modern saat ini, terdapat kebutuhan yang semakin kuat untuk meningkatkan produktivitas sekaligus menjaga keberlanjutan lingkungan. Teknologi informasi serta algoritma machine learning kini banyak dimanfaatkan untuk mendukung proses pengambilan keputusan, termasuk dalam menentukan jenis tanaman yang sesuai berdasarkan parameter lingkungan seperti suhu, kelembapan, pH tanah, dan curah hujan (Bhuyan et al., 2022), (Patel et al., 2023), (Jabed & Murad, 2024), (Paithane, 2023), (M et al., 2024), (Ahmad et al., 2025). Pemanfaatan deteksi objek secara real-time juga memperlihatkan potensi besar

untuk diterapkan pada berbagai sistem cerdas berbasis data di sektor pertanian (Yansah et al., 2025).

Meskipun demikian, praktik di lapangan menunjukkan bahwa banyak petani masih mengandalkan pengalaman dan intuisi ketimbang sistem berbasis data, sehingga menimbulkan kesenjangan antara konsep pertanian presisi dan implementasinya (Elbasi et al., 2023), (Soegoto et al., 2022), (Pramana et al., 2025), (Nova Rahmawati et al., 2025). Penelitian sebelumnya menegaskan bahwa pemanfaatan analitik data dapat membantu proses pengambilan keputusan pada produksi tanaman, meskipun penerapan algoritma machine learning belum sepenuhnya optimal (Adesyahputra & Rachmawati, 2025). Selain itu, sebagian besar penelitian masih berfokus pada pengembangan model atau prototipe bersifat offline dan belum menyediakan platform berbasis web yang dapat digunakan secara langsung oleh petani atau pengguna umum (Kiran et al., 2024). Palupi et al. juga menunjukkan bahwa integrasi variabel lingkungan mampu meningkatkan akurasi klasifikasi tanaman, namun pemanfaatan sistem prediktif berbasis web masih sangat terbatas (Senapaty et al., 2024).

Situasi tersebut mengindikasikan bahwa meskipun potensi machine learning dalam bidang pertanian telah banyak dibuktikan, penerapan solusi yang praktis dan mudah diakses masih minim. Selain itu, penelitian lain menyoroti pentingnya ketersediaan struktur data yang baik serta metode pemodelan yang tepat untuk menghasilkan prediksi yang akurat (Yuan et al., 2024). Berdasarkan konteks tersebut, penelitian ini bertujuan untuk mengembangkan sistem prediksi jenis tanaman berbasis web dengan menggunakan algoritma Decision Tree dan Random Forest, yang mempertimbangkan parameter lingkungan utama seperti suhu, kelembapan, pH tanah, curah hujan, dan intensitas cahaya (Shastri et al., 2025). Sistem ini diharapkan dapat membantu menjembatani kesenjangan antara teori dan praktik, sekaligus mendukung penerapan pertanian berkelanjutan (Baishya & Dutta, 2025), (Alhafiz & Sela, 2025). Hal ini sejalan dengan temuan sebelumnya yang menunjukkan bahwa sistem cerdas seperti irigasi otomatis mampu menurunkan penggunaan air hingga 30% dan meningkatkan efisiensi energi melalui optimalisasi siklus penyiraman (Wahyudi et al., 2025).

## **2. KAJIAN TEORITIS**

Penerapan teknologi informasi dan kecerdasan buatan dalam dunia pertanian semakin berkembang sebagai bagian dari upaya meningkatkan efisiensi dan ketepatan dalam pengambilan keputusan bercocok tanam. Dengan memanfaatkan data lingkungan seperti suhu, kelembapan, pH tanah, dan curah hujan, sistem prediksi mampu memberikan rekomendasi yang lebih terukur dan mengurangi ketergantungan pada intuisi petani. Hal ini sejalan dengan

konsep pertanian presisi yang menekankan penggunaan data sebagai dasar budidaya tanaman yang lebih efektif dan berkelanjutan.

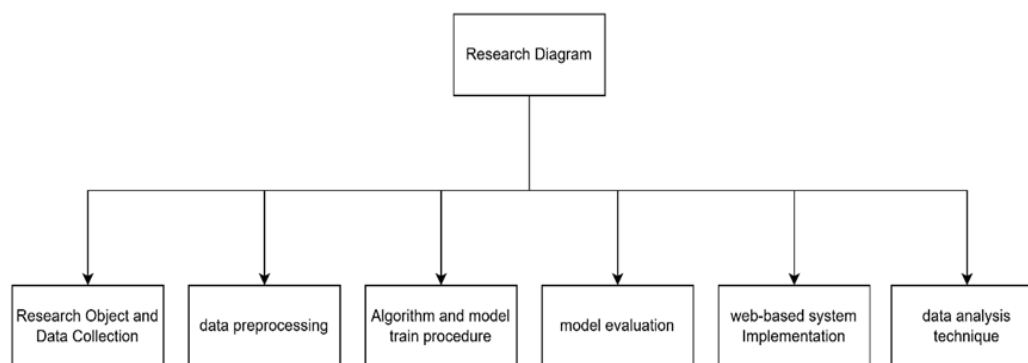
Algoritma Decision Tree merupakan metode klasifikasi yang menyajikan struktur pohon sebagai representasi proses pengambilan keputusan, sehingga hasilnya mudah dipahami oleh pengguna. Sementara itu, Random Forest mengombinasikan sejumlah pohon keputusan melalui pendekatan *ensemble* untuk meningkatkan akurasi dan mengurangi risiko kesalahan model akibat overfitting. Kedua algoritma tersebut telah dibuktikan dalam berbagai penelitian mampu melakukan klasifikasi tanaman dengan performa yang baik pada data lingkungan yang beragam.

Berbagai studi sebelumnya menekankan bahwa masih dibutuhkan solusi yang tidak hanya akurat secara akademik, namun juga mudah diakses oleh pengguna akhir. Sebagian penelitian terdahulu masih sebatas pada pemodelan dan pengujian offline, sehingga belum optimal dalam memberikan manfaat langsung bagi petani dalam pengambilan keputusan sehari-hari. Hal ini menjadi dasar penting bagi penelitian ini untuk menghadirkan sistem prediksi tanaman berbasis web yang responsif, mudah digunakan, dan mampu memberikan hasil secara real-time.

Berdasarkan landasan konsep serta hasil penelitian sebelumnya, penelitian ini mengembangkan sistem prediksi tanaman berbasis algoritma Decision Tree dan Random Forest sebagai upaya menjembatani penerapan machine learning dalam praktik pertanian modern. Secara implisit, penelitian ini mengarahkan bahwa Random Forest berpotensi memberikan performa yang lebih stabil dan akurat dalam memprediksi tanaman dibandingkan Decision Tree ketika diuji pada berbagai variasi parameter lingkungan.

### 3. METODE PENELITIAN

Penelitian ini menggunakan metode eksperimen kuantitatif yang dirancang untuk mengembangkan sekaligus mengevaluasi sistem prediksi tanaman berbasis web. Sistem tersebut memanfaatkan algoritma Decision Tree dan Random Forest guna meningkatkan akurasi prediksi serta kinerja sistem secara keseluruhan (Mancer et al., 2025). Prosedur penelitian dilaksanakan melalui beberapa tahapan berurutan, meliputi pengumpulan data, prapemrosesan data, pengembangan dan pelatihan model, evaluasi kinerja, hingga implementasi akhir system.



**Gambar 1.** Diagram Metodologi Penelitian dan Implementasi Sistem.

Diagram tersebut menggambarkan alur kerja keseluruhan dari metodologi penelitian, yang disusun secara sistematis dimulai dari proses pengumpulan serta pra-pengolahan data. Tahapan ini kemudian dilanjutkan dengan pelatihan model dan evaluasi kinerjanya menggunakan metode machine learning. Pada tahap akhir, hasil model diimplementasikan ke dalam sistem berbasis web dan dianalisis lebih lanjut. Visualisasi ini berfungsi untuk menjelaskan rangkaian proses penelitian secara jelas dan runtut, sehingga mendukung transparansi serta kemudahan reproduksi penelitian.

### **Objek Penelitian dan Pengumpulan Data**

Penelitian ini berfokus pada data lingkungan yang memengaruhi pertumbuhan tanaman, meliputi suhu udara ( $^{\circ}\text{C}$ ), kelembaban (%), pH tanah, curah hujan (mm), serta intensitas cahaya (lux) (Reddy & Neerugatti, 2023). Data diperoleh dari himpunan data pertanian yang tersedia secara publik dan diverifikasi menggunakan data lapangan lokal yang sesuai dengan kondisi agroklimat Indonesia. Seluruh dataset melalui proses pembersihan untuk menghilangkan nilai hilang maupun duplikasi guna memastikan kualitas serta keandalan data.

### **Pra-pemrosesan Data**

Sebelum proses pelatihan dimulai, dataset terlebih dahulu melalui tahap normalisasi dan pengkodean. Data numerik distandardisasi agar berada dalam rentang nilai yang seragam, sedangkan variabel kategorikal termasuk jenis tanaman diubah ke bentuk numerik menggunakan teknik label encoding. Langkah pra-pemrosesan ini penting untuk memastikan seluruh fitur dapat dibaca, dipahami, dan diolah secara optimal oleh model machine learning.

### **Prosedur Algoritma dan Pelatihan Model**

Dua algoritma utama digunakan dalam penelitian ini, yaitu Decision Tree dan Random Forest. Decision Tree membangun struktur berbentuk pohon secara hierarkis dengan melakukan pemilahan atribut berdasarkan nilai information gain tertinggi pada setiap tahap pemisahan (H R & S, 2025), (Qanith & Khilar, 2025). Random Forest menghasilkan

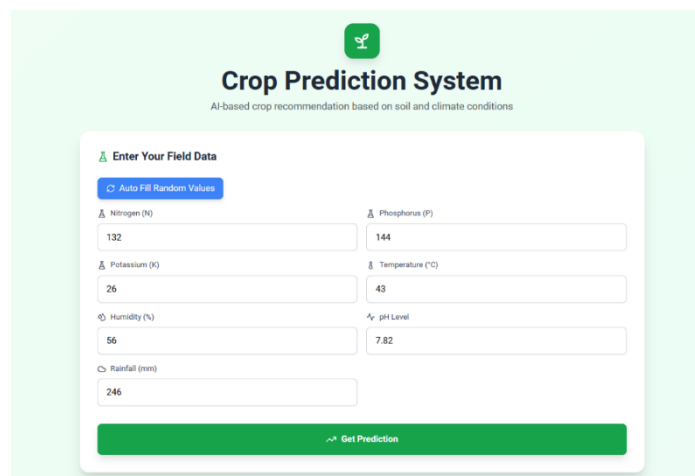
sekumpulan pohon keputusan yang dibentuk secara acak, di mana hasil prediksi akhir diperoleh melalui mekanisme voting mayoritas dari seluruh pohon yang terbentuk (Brandt et al., 2024). Dataset pada penelitian ini dibagi menjadi dua bagian, yaitu 80% sebagai data latih dan 20% sebagai data uji, menggunakan teknik train-test split untuk memastikan evaluasi model yang proporsional. Proses pelatihan model dilakukan menggunakan pustaka Scikit-Learn pada bahasa pemrograman Python. Sejumlah hiperparameter meliputi `max_depth`, `n_estimators`, dan `criterion` dioptimalkan melalui metode grid search untuk memperoleh konfigurasi model dengan performa paling optimal.

### Evaluasi Model

Evaluasi kinerja model dilakukan dengan menggunakan empat metrik utama, yaitu akurasi, presisi, recall, dan skor F1 (Addu et al., 2024). Selain itu, confusion matrix juga dimanfaatkan untuk menilai sebaran klasifikasi benar maupun salah pada setiap jenis tanaman. Model dengan nilai F1-score tertinggi ditetapkan sebagai model paling optimal, karena metrik ini memberikan keseimbangan yang lebih menyeluruh antara presisi dan recall dalam mengukur performa model.

### Implementasi Sistem Berbasis Web

Model paling efektif kemudian diintegrasikan ke dalam aplikasi berbasis web. Pada sisi backend, sistem dikembangkan menggunakan framework Flask, sedangkan antarmuka pengguna dirancang dengan HTML, CSS, dan JavaScript. Melalui aplikasi ini, pengguna dapat memasukkan parameter lingkungan secara langsung melalui formulir web, dan sistem akan menghasilkan prediksi jenis tanaman secara real-time berdasarkan model yang telah dilatih.

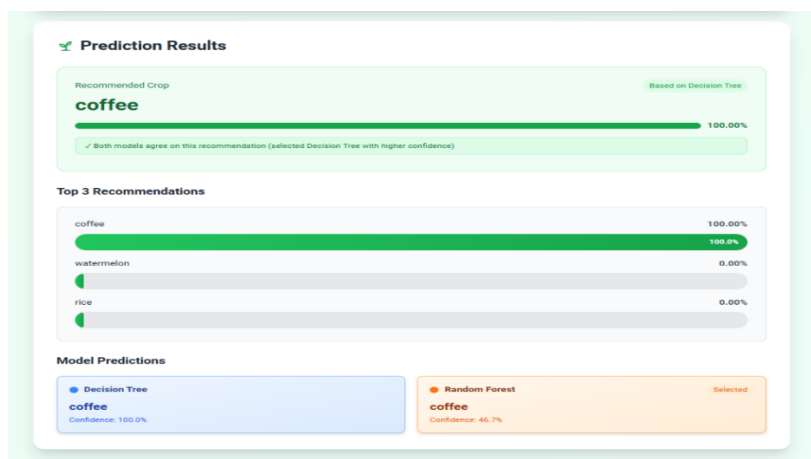


The screenshot displays the 'Crop Prediction System' web interface. At the top, there is a green header with a leaf icon and the title 'Crop Prediction System' followed by the subtitle 'AI-based crop recommendation based on soil and climate conditions'. Below this is a white form titled 'Enter Your Field Data'. The form contains several input fields with labels and icons: Nitrogen (N) with a value of 132, Phosphorus (P) with a value of 144, Potassium (K) with a value of 26, Temperature (°C) with a value of 43, Humidity (%) with a value of 56, pH Level with a value of 7.82, and Rainfall (mm) with a value of 246. There is a blue button labeled 'Auto Fill Random Values' and a green button labeled 'Get Prediction' at the bottom of the form.

**Gambar 2.** Antarmuka Input Data Pertanian.

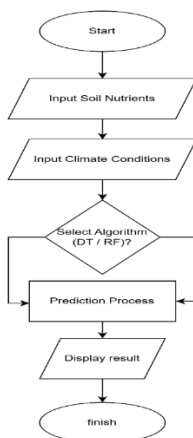
Bagian ini menampilkan antarmuka formulir input dari *Crop Prediction System*. Pada tahap ini, pengguna memasukkan berbagai data penting terkait unsur hara tanah (N, P, K, dan pH) serta kondisi iklim (suhu, kelembapan, dan curah hujan) ke dalam kolom yang tersedia.

Sistem kemudian mengolah seluruh informasi tersebut untuk menghasilkan rekomendasi tanaman yang paling sesuai.



**Gambar 3.** Antarmuka Hasil Prediksi Tanaman.

Antarmuka ini menampilkan formulir input untuk *Crop Prediction System*. Pengguna memasukkan data penting terkait kandungan hara tanah (N, P, K, dan pH) serta kondisi iklim (suhu, kelembapan, dan curah hujan) ke dalam kolom yang telah disediakan. Selanjutnya, sistem mengolah data tersebut untuk menghasilkan rekomendasi tanaman yang paling sesuai berdasarkan parameter yang telah dianalisis.



**Gambar 4.** Diagram Alur Prediksi Hasil Panen Menggunakan Decision Tree atau Random Forest.

Bagan alir tersebut menggambarkan alur kerja dari sistem prediksi tanaman. Proses dimulai dengan pemasukan data nutrisi tanah serta kondisi iklim. Selanjutnya, sistem memilih algoritma prediksi yang sesuai baik Decision Tree (DT) maupun Random Forest (RF) untuk melakukan analisis. Pada tahap akhir, sistem menampilkan rekomendasi jenis tanaman kepada pengguna berdasarkan hasil prediksi yang telah dihasilkan.

## Teknik Analisis Data

Hasil prediksi dan kinerja model dianalisis secara kuantitatif melalui perbandingan nilai akurasi dan F1-score pada setiap algoritma. Analisis ini bertujuan untuk mengidentifikasi algoritma yang memberikan performa terbaik berdasarkan kondisi data lingkungan di Indonesia, serta menilai sejauh mana setiap parameter lingkungan berkontribusi terhadap keluaran prediksi.

**Tabel 1.** Ringkasan Tahapan Penelitian.

Tahap Kegiatan Utama	Tujuan
1 Pengumpulan Data	Kumpulkan data lingkungan dan tanaman
2 Pemrosesan Awal	Siapkan data untuk machine learning
3 Pelatihan Model	Melatih Decision Tree and Random Forest algorithms
4 Evaluasi Model	Menilai kinerja model menggunakan evaluation metrics
5 Implementasi Web	Mengembangkan system prediksi tanamann berbasis web
6 Analisis Data	Menafsirkan hasil dan menarik kesimpulan

## Rumus

Akurasi merupakan persentase prediksi yang benar yang dihasilkan oleh suatu model dibandingkan dengan seluruh prediksi yang dilakukan. Secara formal, akurasi dihitung dengan membagi jumlah kasus yang berhasil diidentifikasi dengan tepat—yakni True Positive (TP) dan True Negative (TN)—dengan total keseluruhan data ( $TP + TN + FP + FN$ ). Metode ini memberikan gambaran umum mengenai tingkat ketepatan model dalam melakukan klasifikasi. Namun demikian, penggunaan akurasi sebagai satu-satunya indikator dapat menimbulkan bias, terutama pada kondisi dataset yang tidak seimbang. Pada situasi tersebut, model dapat terlihat memiliki akurasi tinggi hanya karena lebih sering memprediksi kelas mayoritas, tanpa benar-benar mampu membedakan seluruh kelas secara efektif.

Table 1. Model Evaluation: Accuracy & F1 Score

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2)$$

Sebaliknya, F1-Score merupakan metrik evaluasi yang menggabungkan dua komponen utama dalam efektivitas klasifikasi, yaitu Precision dan Recall. Nilai ini diperoleh melalui perhitungan mean harmonik antara precision dan recall, sehingga memberikan representasi yang lebih seimbang dari keduanya. Metrik ini berfungsi untuk menjaga proporsi yang adil antara dua parameter tersebut, sehingga sangat sesuai digunakan pada kondisi dengan distribusi kelas yang tidak seimbang. Dengan demikian, F1-Score menjadi indikator yang lebih objektif

dalam menilai kinerja model, terutama ketika kesalahan pada satu kelas memiliki dampak yang lebih besar dibandingkan kelas lainnya.

#### 4. HASIL DAN PEMBAHASAN








##### Temuan Penelitian

Penelitian ini menggunakan data lingkungan dan nutrisi yang dianalisis melalui algoritma Decision Tree (DT) dan Random Forest (RF) dengan memanfaatkan platform Orange Data Mining. Fitur-fitur yang diamati meliputi kelembapan, curah hujan, pH tanah, serta kandungan nitrogen (N), fosfor (P), dan kalium (K). Dataset dibagi menggunakan teknik 10-fold cross-validation, di mana pada setiap iterasi, sembilan subset digunakan untuk pelatihan model, sedangkan satu subset sisanya digunakan untuk pengujian. Proses ini diulang sebanyak sepuluh kali, dan metrik kinerja akhir diperoleh dengan mengambil rata-rata hasil dari seluruh iterasi.

##### Analisis Pentingnya Fitur (Hasil Peringkat)

Hasil Rank Widget menggunakan metode Information Gain, Gain Ratio, dan Gini Decrease menunjukkan bahwa kelembapan memiliki kontribusi terbesar terhadap akurasi klasifikasi tanaman, dengan nilai information gain sebesar 1,567, diikuti oleh curah hujan (1,504), kalium (K) (1,356), dan fosfor (P) (1,261). Sebaliknya, pH tanah memiliki pengaruh paling rendah terhadap klasifikasi, dengan nilai information gain sebesar 0,640 dan gain ratio sebesar 0,320.

Hasil pengujian menunjukkan bahwa fitur kelembapan dan curah hujan memberikan kontribusi paling besar terhadap hasil klasifikasi tanaman.

		#	Info. gain	Gain ratio
1	 humidity		1.567	0.783
2	 rainfall		1.504	0.752
3	 K		1.356	0.678
4	 P		1.261	0.630
5	 N		0.960	0.480
6	 temperature		0.833	0.416
7	 ph	nan	0.640	0.320

**Gambar 5.** Grafik Nilai Peningkatan Informasi untuk Setiap Fitur.

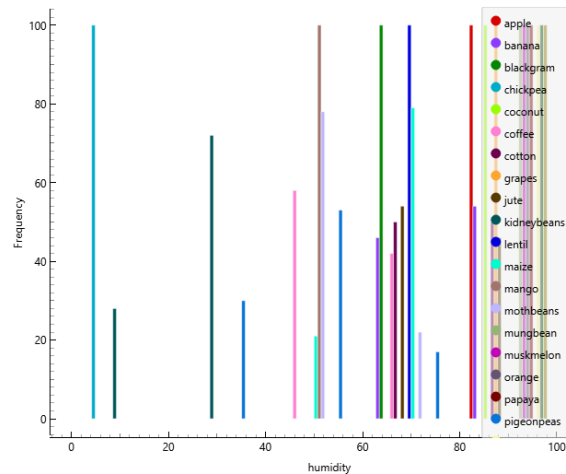
Hasil ini menunjukkan bahwa kelembapan udara dan curah hujan merupakan variabel lingkungan yang paling berpengaruh dalam menentukan jenis tanaman. Semakin tinggi nilai



information gain, semakin besar kontribusi fitur tersebut dalam mengurangi entropi serta meningkatkan kepastian klasifikasi.

### ***Analisis Distribusi Data***

Hasil distribusi data menunjukkan bahwa tanaman yang membutuhkan tingkat nitrogen sedang hingga tinggi cenderung mendominasi kategori tropis dengan kelembapan tinggi. Sebagai contoh, pada rentang nitrogen  $50 \leq N < 100$ , tanaman seperti jute (16,33%), mangga (10,46%), dan mungbean (8,08%) muncul paling sering.



**Gambar 6.** Distribusi Nilai Nitrogen (N) pada Jenis Tanaman.

Hal ini menunjukkan adanya korelasi yang kuat antara kandungan nitrogen dan kelembapan lingkungan dalam menentukan pertumbuhan tanaman yang optimal.

### ***Evaluasi Model dan Perbandingan Kinerja***

Kinerja model dievaluasi menggunakan berbagai metrik penilaian, yaitu Area Under the Curve (AUC), Akurasi Klasifikasi (CA), Presisi, Recall, dan F1-score.

Model	AUC	CA	F1	Prec	Recall	LogLoss
random forest	1.000	0.995	0.995	0.996	0.995	0.048
decision tree	0.991	0.977	0.977	0.977	0.977	0.615

**Gambar 7.** Perbandingan Nilai Akurasi, Presisi, dan F1-Score antara Algoritma Decision Tree dan Random Forest.

Dataset dibagi menjadi 10 subset, di mana 9 subset digunakan untuk pelatihan dan 1 subset digunakan untuk pengujian secara bergantian. Hasil akhir diperoleh dari rata-rata dari 10 percobaan tersebut..

### ***AUC (Area Under the Curve)***

Fungsi = mengevaluasi seberapa efektif model dalam membedakan berbagai kelas pada rentang 0–1. Semakin mendekati nilai 1, semakin baik hasil yang diperoleh.

RF = Nilai 1,000 dan DT sebesar 0,991 menunjukkan keduanya sangat baik, namun RF menunjukkan performa yang lebih sempurna.

#### *CA (Classification Accuracy)*

Fungsi = Akurasi total (persentase prediksi yang benar).

RF = 0.995 → 99.5% Akurat; DT = 0.977 → 97.7%

#### *F1*

Fungsi = Rata-rata harmonik dari presisi dan recall; digunakan untuk menilai keseimbangan kinerja model.

RF = 0.995, DT = 0.977 → Random Forest memiliki kestabilan dan tingkat presisi yang lebih tinggi.

#### *Prec (Precision)*

Fungsi: Dari seluruh prediksi positif, berapa banyak yang terbukti benar.

RF = 0.996 → sangat tepat.

#### *Recall (Sensitivity)*

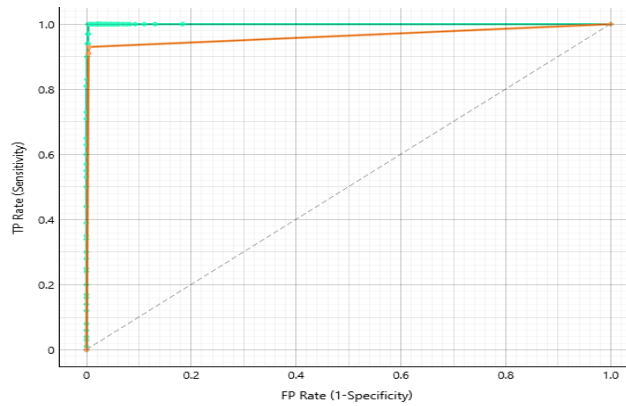
Fungsi: Dari seluruh kasus positif yang sebenarnya, berapa banyak yang berhasil dideteksi oleh model?

RF = 0.995 → Hampir seluruh kasus berhasil dideteksi dengan benar.

Nilai AUC sebesar 1,000 pada model Random Forest menunjukkan kemampuan yang sangat baik dalam membedakan berbagai kategori tanaman secara akurat, sedangkan Decision Tree juga menampilkan kinerja yang kuat dengan AUC = 0,991. Secara keseluruhan, algoritma Random Forest menunjukkan performa yang lebih unggul dibandingkan Decision Tree, dengan stabilitas, presisi, dan kemampuan generalisasi yang lebih tinggi.

#### ***Analisis ROC dan Confusion Matriks***

Analisis ROC (Receiver Operating Characteristic) menunjukkan bahwa model Random Forest memiliki kinerja yang lebih unggul, ditandai oleh kurvanya yang berada lebih dekat ke sudut kiri atas. Posisi ini mencerminkan tingginya tingkat positif sejati (True Positive Rate/TPR) bersamaan dengan rendahnya tingkat positif palsu (False Positive Rate/FPR), yang menegaskan kemampuan diskriminatif model tersebut secara kuat.



**Gambar 8.** Kurva ROC pada Algoritma Decision Tree dan Random Forest.

TPR (True Positive Rate) atau Sensitivitas: Mengukur seberapa banyak kasus positif yang berhasil terdeteksi dengan benar.

FPR (False Positive Rate) atau 1 Spesifisitas: Mengukur seberapa banyak kasus negatif yang salah diklasifikasikan sebagai positif.

Semakin mendekati sisi kiri atas (TPR lebih tinggi, FPR lebih rendah), semakin baik kinerja model. Hal ini menunjukkan bahwa model Random Forest lebih efektif dalam mengidentifikasi jenis tanaman dengan tingkat kesalahan klasifikasi yang minimal.

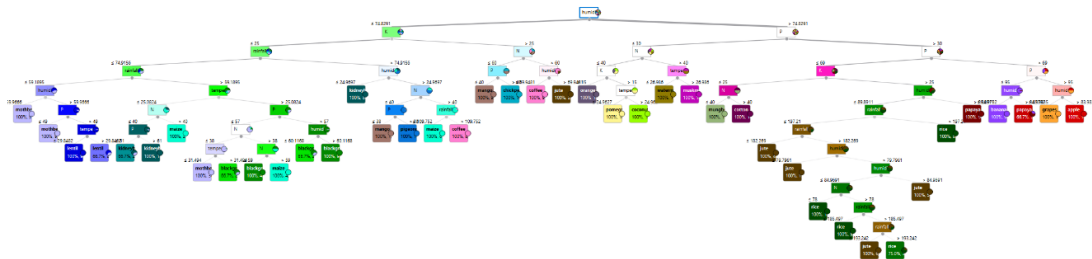
	apple	banana	blackgr...	chickpea	coconut	coffee	cotton	graj
apple	100	0	0	0	0	0	0	0
banana	0	100	0	0	0	0	0	0
blackgr...	0	0	99	0	0	0	0	0
chickpea	0	0	0	100	0	0	0	0
coconut	0	0	0	0	100	0	0	0
coffee	0	0	0	0	0	100	0	0

**Gambar 9.** Matriks Kebingungan.

Fungsinya adalah untuk mengukur seberapa akurat model dalam memprediksi setiap kelas (label tanaman) tidak hanya akurasi keseluruhan, tetapi juga untuk masing-masing kelas secara individu. Matriks kebingungan (Confusion Matrix) juga menunjukkan bahwa hampir semua kelas tanaman berhasil diprediksi dengan benar, sehingga menegaskan akurasi baik secara keseluruhan maupun per kelas.

### **Visualisasi Decision Tree**

Hasil visualisasi menunjukkan bahwa kelembapan dan curah hujan sering muncul pada simpul teratas (simpul akar) dalam struktur pohon keputusan. Temuan ini memperkuat hasil analisis peringkat sebelumnya, yang menegaskan bahwa kedua variabel tersebut memiliki peran paling penting dalam membedakan kategori tanaman.



**Gambar 10.** Visualisasi Pohon Keputusan.

Fitur 'kelembapan' dan 'curah hujan' paling sering muncul pada simpul awal → hal ini menegaskan hasil perhitungan Information Gain sebelumnya bahwa kedua fitur ini merupakan faktor penentu utama.

**Hasil**

Penelitian ini bertujuan untuk memprediksi jenis tanaman berdasarkan parameter lingkungan dengan memanfaatkan dua algoritma pembelajaran mesin: Decision Tree (DT) dan Random Forest (RF). Temuan penelitian dirangkum dan dibahas secara mendetail, didukung oleh representasi data baik secara visual maupun kuantitatif.

**Tinjauan Umum Hasil Eksperimen**

Dataset diproses menggunakan platform Orange Data Mining dan dibagi menjadi sepuluh subset melalui 10-fold cross-validation, sehingga kedua model dapat dilatih dan diuji secara merata. Hasil analisis menunjukkan bahwa seluruh model menunjukkan kinerja yang sangat baik, dengan algoritma Random Forest mencapai performa keseluruhan tertinggi.

**Tabel 2.** Ringkasan Metrik Evaluasi Model.

Metric	Decision Tree (DT)	Random Forest (RF)
AUC	0.991	<b>1.000</b>
Accuracy	0.977	<b>0.995</b>
F1-Score	0.977	<b>0.995</b>
Precision	0.974	<b>0.996</b>
Recall	0.976	<b>0.995</b>

**Penjelasan:**

Hasil pada Tabel 1 menunjukkan bahwa model Random Forest memiliki kinerja lebih baik dibandingkan Decision Tree pada seluruh metrik evaluasi. Skor AUC sebesar 1,000 mengindikasikan bahwa model Random Forest mampu membedakan kelas tanaman dengan sempurna. Tingkat akurasi sebesar 99,5% juga menegaskan tingkat keandalan yang tinggi dan minimnya kesalahan klasifikasi, sehingga memperlihatkan kekokohan model serta kemampuan generalisasi yang kuat.

### *Analisis Pentingnya Fitur*

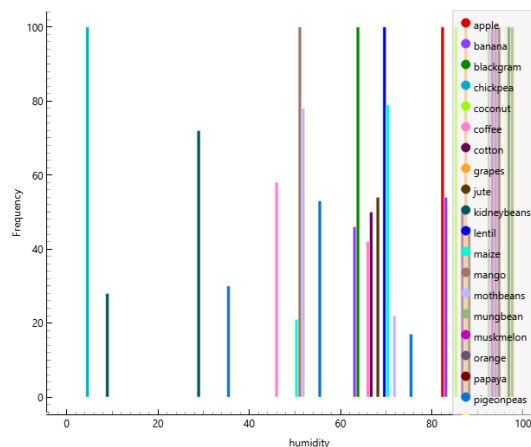
Untuk mengetahui fitur-fitur yang paling berpengaruh terhadap hasil prediksi, dilakukan perankingan fitur menggunakan metode Information Gain, Gain Ratio, dan Gini Decrease.

		#	Info. gain	Gain ratio
1	<b>N</b> humidity		1.567	0.783
2	<b>N</b> rainfall		1.504	0.752
3	<b>N</b> K		1.356	0.678
4	<b>N</b> P		1.261	0.630
5	<b>N</b> N		0.960	0.480
6	<b>N</b> temperature		0.833	0.416
7	<b>N</b> ph	nan	0.640	0.320

**Gambar 11.** Nilai Information Gain untuk Setiap Fitur.

Hasil analisis menunjukkan bahwa kelembapan (1,567) dan curah hujan (1,504) merupakan dua parameter yang paling dominan dalam mempengaruhi klasifikasi tanaman, diikuti oleh kadar kalium (1,356) dan fosfor (1,261). Sementara itu, pH tanah (0,640) memberikan kontribusi paling rendah. Temuan ini mengindikasikan bahwa faktor iklim dan kelembapan tanah memiliki peran yang lebih menentukan dalam membedakan jenis tanaman dibandingkan dengan tingkat keasaman tanah.

### *Pola Sebaran Tanaman*



**Gambar 12.** Distribusi Nilai Nitrogen (N) Berdasarkan Jenis Tanaman.

Distribusi fitur nitrogen menunjukkan bahwa tanaman dengan kebutuhan nitrogen sedang hingga tinggi didominasi oleh spesies tropis yang memerlukan kelembapan tinggi. Sebagai contoh, tanaman seperti jute (16,33%), mangga (10,46%), dan kacang hijau (8,08%) menempati porsi terbesar dalam dataset pada rentang nitrogen  $50 \leq N < 100$ . Tren ini menggambarkan bahwa konsentrasi nitrogen memiliki keterkaitan erat dengan kemampuan adaptasi tanaman di lingkungan yang lembap.

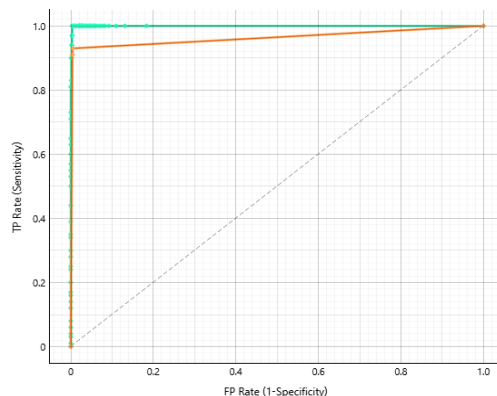
### Perbandingann kinerja model

Model	AUC	CA	F1	Prec	Recall	LogLoss
random forest	1.000	0.995	0.995	0.996	0.995	0.048
decision tree	0.991	0.977	0.977	0.977	0.977	0.615

**Gambar 13.** Perbandingan Akurasi, Presisi, Recall, dan Skor F1.

Kedua algoritma menunjukkan akurasi yang tinggi; namun, Random Forest secara konsisten menampilkan kinerja yang lebih baik dibandingkan Decision Tree pada seluruh metrik yang digunakan. Perbedaan kinerja yang relatif kecil menunjukkan bahwa kedua model dapat diandalkan untuk prediksi tanaman, meskipun Random Forest menawarkan konsistensi yang lebih tinggi berkat pendekatan ensemble nya yang mampu mengurangi overfitting. Hipotesis yang diajukan dalam penelitian ini bahwa Random Forest akan memberikan prediksi yang lebih akurat dan stabil dibandingkan Decision Tree terbukti benar berdasarkan hasil eksperimen.

### Kurva ROC dan Perilaku Klasifikasi



**Gambar 14.** Kurva ROC pada Algoritma Decision Tree dan Random Forest.

Analisis ROC menunjukkan bahwa model Random Forest menghasilkan kurva yang berada di dekat sudut kiri atas plot, yang mencerminkan tingkat true positive (TPR) yang tinggi sekaligus tingkat false positive (FPR) yang rendah, sehingga menunjukkan kinerja klasifikasi yang kuat. Hal ini menegaskan bahwa model mampu membedakan klasifikasi positif (benar) dari negatif dengan kesalahan yang minimal.

Selain itu, analisis Confusion Matrix menunjukkan bahwa sebagian besar jenis tanaman diklasifikasikan dengan benar, yang semakin menguatkan kemampuan model dalam mencapai presisi tinggi serta kemampuan generalisasi yang baik.

### Interpretasi Hasil

Berdasarkan temuan secara keseluruhan, beberapa kesimpulan utama dapat ditarik sebagai berikut: Variabel lingkungan yang dianalisis menunjukkan bahwa kelembapan dan

curah hujan merupakan faktor paling signifikan dalam memprediksi jenis tanaman, menegaskan peran dominan faktor iklim dalam model klasifikasi pertanian. Algoritma Random Forest menunjukkan kinerja yang lebih unggul dibandingkan Decision Tree pada seluruh metrik evaluasi, mendukung pemahaman bahwa metode ensemble umumnya memberikan akurasi prediksi yang lebih tinggi serta stabilitas model yang lebih baik. Platform berbasis web yang dikembangkan berhasil mengintegrasikan algoritma Random Forest, memungkinkan prediksi jenis tanaman secara real-time melalui antarmuka browser yang intuitif dan mudah diakses. Hasil penelitian ini menunjukkan bahwa penerapan machine learning dalam pertanian presisi sangat memungkinkan, khususnya untuk lingkungan tropis seperti Indonesia.

### **Diskusi**

Hasil penelitian menunjukkan bahwa penerapan algoritma Decision Tree dan Random Forest dalam sistem berbasis web mampu mencapai akurasi tinggi dalam memprediksi jenis tanaman berdasarkan parameter lingkungan. Model Decision Tree menawarkan keunggulan dalam hal interpretabilitas, sehingga pengguna dapat dengan mudah memahami pola pengambilan keputusan yang dihasilkan dari data. Sementara itu, model Random Forest menunjukkan stabilitas dan akurasi yang lebih tinggi dengan meminimalkan risiko overfitting melalui kombinasi beberapa pohon keputusan.

Temuan ini memperkuat pemahaman bahwa pendekatan berbasis data dalam analisis lingkungan dapat secara efektif mendukung identifikasi jenis tanaman. Integrasi sistem ini ke dalam platform berbasis web juga memungkinkan akses yang lebih luas serta penerapan secara real-time dalam berbagai kondisi lapangan. Oleh karena itu, sistem ini berpotensi menjadi alat yang efektif untuk mendukung proses pengambilan keputusan bagi peneliti, petani, dan pembuat kebijakan dalam menentukan jenis tanaman yang sesuai berdasarkan kondisi lingkungan tertentu.

Secara teoretis, penelitian ini menekankan peran penting pemodelan berbasis data dalam mendukung proses pengambilan keputusan di bidang pertanian modern. Pendekatan berbasis pohon keputusan menyediakan kerangka kerja yang adaptif dan dapat diperluas untuk analisis prediktif di sektor lain yang bergantung pada data lingkungan. Secara praktis, hasil penelitian ini dapat digunakan untuk mengoptimalkan sistem rekomendasi budidaya tanaman, sehingga penggunaan sumber daya menjadi lebih efisien dan produktivitas meningkat secara keseluruhan.

Lebih lanjut, hasil penelitian ini membuka peluang pengembangan sistem yang lebih cerdas melalui integrasi model prediktif dengan data sensor lingkungan secara real-time. Oleh karena itu, penelitian ini tidak hanya memperkuat dasar ilmiah penerapan kecerdasan buatan

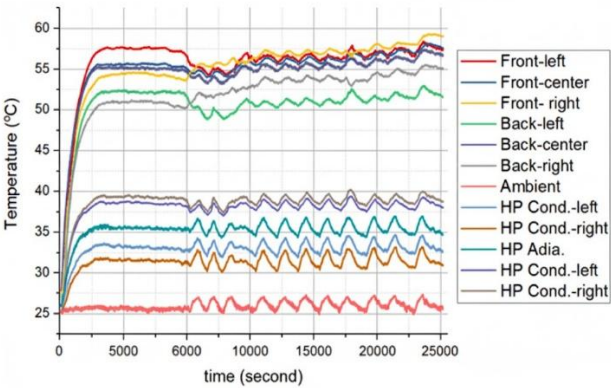
di bidang pertanian, tetapi juga berperan penting dalam mendorong kemajuan teknologi digital untuk mendukung keberlanjutan dan meningkatkan ketahanan pangan di masa depan.

Tabel Dan Gambar

Tabel 2. Konsumsi daya, biaya, dan waktu.

Bulan	Konsumsi (kWh)	Biaya (USD)
Januari	120	10.50
Februari	100	8.75
Maret	150	13.00
April	130	11.38
Mei	160	14.00

Menyajikan hubungan antara konsumsi daya (kWh) dan biaya operasional (USD) selama periode lima bulan. Data menunjukkan adanya peningkatan bertahap pada penggunaan energi dari Januari hingga Mei. Konsumsi terendah tercatat pada bulan Februari sebesar 100 kWh dengan biaya USD 8,75, sedangkan konsumsi tertinggi terjadi pada bulan Mei sebesar 160 kWh dengan biaya USD 14,00. Tren kenaikan ini menunjukkan adanya peningkatan permintaan operasional atau beban kerja sistem yang lebih tinggi pada bulan-bulan terakhir periode pengamatan.



Gambar 15. Karakteristik Suhu.

Grafik ini menyajikan perbandingan throughput (per detik) dari tiga sistem Sistem A, Sistem B, dan Sistem C yang telah dioptimalkan dihitung selama 1.000 detik. Secara keseluruhan, grafik ini menggambarkan laju pertumbuhan produktivitas masing-masing sistem seiring waktu. Sistem A (garis biru) digunakan sebagai acuan, menunjukkan peningkatan yang konsisten. Sistem B (garis oranye putus-putus) memperlihatkan pertumbuhan throughput tercepat, meskipun dimulai dari level yang lebih rendah, namun akhirnya mencapai throughput tertinggi pada akhir periode pengujian. Sebaliknya, Sistem C yang dioptimalkan (garis hijau putus-putus), meskipun memiliki label “dioptimalkan”, menunjukkan throughput terendah secara keseluruhan dan pertumbuhan kinerja paling lambat selama 1.000 detik pengukuran.



## 5. KESIMPULAN DAN SARAN

Selain itu, pengembangan platform prediksi berbasis web memungkinkan pengguna untuk memasukkan parameter lingkungan secara langsung dan memperoleh rekomendasi tanaman yang sesuai secara real-time. Evaluasi kinerja menunjukkan bahwa Random Forest unggul dibandingkan Decision Tree dalam hal akurasi, presisi, dan kestabilan, dengan nilai AUC sebesar 1,000 dan akurasi mencapai 99,5%. Analisis fitur mengungkapkan bahwa kelembaban dan curah hujan merupakan faktor yang paling berpengaruh dalam klasifikasi tanaman tropis, sedangkan pH tanah memiliki pengaruh yang paling rendah. Temuan ini menegaskan pentingnya kondisi iklim dan kelembaban tanah dalam menentukan pertumbuhan tanaman yang optimal.

Selain itu, pengembangan sistem prediksi berbasis web memberikan kemudahan bagi pengguna untuk memasukkan data lingkungan secara langsung dan menerima rekomendasi tanaman secara instan. Integrasi model Random Forest pada platform web ini menunjukkan bahwa machine learning dapat mendukung pertanian presisi secara efektif, meningkatkan efisiensi penggunaan sumber daya, serta memfasilitasi pengambilan keputusan berbasis data. Sistem ini menjadi jembatan antara penelitian akademik dan praktik pertanian nyata, memberikan alat yang ramah pengguna bagi petani untuk meningkatkan produktivitas secara berkelanjutan.

Studi lanjutan disarankan untuk mempertimbangkan integrasi data sensor lingkungan secara real-time dan memperluas cakupan faktor yang dianalisis, termasuk intensitas cahaya, kelembaban tanah pada berbagai kedalaman, serta keberadaan hama. Selain itu, pengujian sistem pada berbagai wilayah agroklimat yang berbeda dapat meningkatkan akurasi dan ketahanan model, mendukung adopsi yang lebih luas oleh petani, serta memfasilitasi praktik pertanian yang lebih efisien dan berkelanjutan.

## DAFTAR PUSTAKA

- Addu, S., Sheelam, S., Mekala, S., Sulthana, N., Mekala, L., & Alsalami, Z. (2024). Assessing Environmental Impact: Machine Learning for Crop Yield Prediction. *E3S Web of Conferences*, 529, 3008.
- Adesyahputra, M. K., & Rachmawati, E. P. (2025). *Analysis of corn production in Indonesia using business intelligence technology based on Power BI*. 15(1), 21–31.
- Ahmad, I., Informasi, S., Teknokrat, U., & Lampung, B. (2025). *Persentase Penyakit Alpukat*. 5(1), 237–243.
- Alhafiz, A. D., & Sela, E. I. (2025). Aplikasi Mobile Untuk Konsultasi Petani Dalam Mendukung Pertanian Digital. *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(1), 9–14. <https://doi.org/10.52436/1.jpti.453>

- Baishya, M., & Dutta, L. (2025). Tiny ML Based Crop Recommendation System for Precision Agriculture 5.0. *Smart Agricultural Technology*, 101247.
- Brandt, P., Beyer, F., Borrmann, P., Möller, M., & Gerighausen, H. (2024). Ensemble learning-based crop yield estimation: a scalable approach for supporting agricultural statistics. *GIScience & Remote Sensing*, 61(1), 2367808.
- Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., Shdefat, A., & Saker, L. (2023). Crop prediction model using machine learning algorithms. *Applied Sciences*, 13(16), 9288.
- H R, M. D., & S, S. T. (2025). International Journal of Research Publication and Reviews Employee Performance Prediction Using Machine Learning. *International Journal of Research Publication and Reviews*, 6(6), 8659–8661. [www.ijrpr.com](http://www.ijrpr.com)
- Kiran, P. S., Abhinaya, G., Sruti, S., & Padhy, N. (2024). A machine learning-enabled system for crop recommendation. *Engineering Proceedings*, 67(1), 51.
- M, M. H., Jogi, D. K., Dalavai M, A. S., Dalli, M., & M Dhanush Jogi K Adarsha Dalavai S M Manvith Dalli, M. H. (2024). *A Web Based Crop Recommendation System Using Various Machine Learning Algorithms*. 9(2), 160–165. [www.ijnrd.org](http://www.ijnrd.org)
- Mancer, M., Terrissa, S. L., & Ayad, S. (2025). *A Systematic Literature Review of Crop Recommendation Systems for Agriculture 4.0*.
- Nova Rahmawati, E., Pinandita, T., Ayu Fitriani, M., & Ambar Pambudi, E. (2025). Deteksi Penyakit Daun Tomat Real-Time pada Platform Android Berbasis Convolutional Neural Network. *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(8), 2402–2415. <https://doi.org/10.52436/1.jpti.964>
- Palupi, S., Gunawan, G., Kusdyawati, R., Hardi, R., & Zabrina, R. (2024). Predicting financial default risks: A machine learning approach using smartphone data. *MATRIX: Jurnal Manajemen Teknologi Dan Informatika*, 14(3), 107–118. <https://doi.org/10.31940/matrix.v14i3.107-118>
- Pramana, A., Dalimunthe, E. R., Komputer, T., Teknik, F., Indonesia, U. T., Elektro, T., Teknik, F., Indonesia, U. T., Informasi, S., Teknik, F., & Indonesia, U. T. (2025). *Teknologi Pemberian Nutrisi Ikan Lele dan Tanaman Kangkung pada Sistem Aquaponik Menggunakan Teknologi IoT Technology for Providing Nutrition to Catfish and Water Spinach in Aquaponic Systems Using IoT*. 5(3), 813–829.
- Qanith, M., & Khilar, R. (2025). Improving accuracy for recommending the appropriate crop using random forest in comparison with decision tree. *AIP Conference Proceedings*, 3252(1), 20001.
- Reddy, D. M. S., & Neerugatti, U. R. (2023). A comparative analysis of machine learning models for crop recommendation in India. *Revue d'Intelligence Artificielle*, 37(4), 1081.
- Senapaty, M. K., Ray, A., & Padhy, N. (2024). A decision support system for crop recommendation using machine learning classification algorithms. *Agriculture*, 14(8), 1256.
- Shastri, S., Kumar, S., Mansotra, V., & Salgotra, R. (2025). Advancing crop recommendation system with supervised machine learning and explainable artificial intelligence. *Scientific Reports*, 15(1), 25498.
- Soegoto, E. S., Warlina, L., Supatmi, S., Rafdhi, A. A., Jumansyah, R., & Saputra, H. (2022).

SITAMPAN: Mobile application for planting and harvesting of horticultural crops in Garut Regency. *MATRIX: Jurnal Manajemen Teknologi Dan Informatika*, 12(3), 123–136. <https://doi.org/10.31940/matrix.v12i3.123-136>

- Wahyudi, W., Pradana, A. I., & Permatasari, H. (2025). Implementasi Sistem Irigasi Otomatis Berbasis IoT untuk Pertanian Greenhouse. *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(2), 435–446. <https://doi.org/10.52436/1.jpti.656>
- Yansah, M. K., Dijaya, R., Setiawan, H., & Sumarno, S. (2025). Humanoid object detection moving in open space using YOLOv8. *MATRIX: Jurnal Manajemen Teknologi Dan Informatika*, 15(2), 60–71. <https://doi.org/10.31940/matrix.v15i2.60-71>
- Yuan, J., Zhang, Y., Zheng, Z., Yao, W., Wang, W., & Guo, L. (2024). Grain crop yield prediction using machine learning based on UAV remote sensing: A systematic literature review. *Drones*, 8(10), 559.