

Penerapan Aplikasi Sistem Rekomendasi Pemilihan Varietas Tanaman Berbasis Parameter Kelembaban Tanah, pH, dan Kandungan Nutrisi Pada Gapoktan Sri Karya Menggunakan Algoritma *Random Forest*

Junus Sinuraya¹, Weno Syechu², Wiwin Sry Adinda Banjarnahor³, Muhammad Riki Atsauri⁴

¹²³⁴⁵Teknik Komputer, Politeknik Negeri Medan

Email: ¹junussinuraya@polmed.ac.id, ²wenosyechu@polmed.ac.id, ³wiwinbanjarnahor@polmed.ac.id,
⁴riki@polmed.ac.id

Email Penulis Korespondensi: ¹junussinuraya@polmed.ac.id

Abstrak

Pemilihan varietas tanaman yang sesuai dengan kondisi tanah lahan pertanian merupakan faktor kritis dalam meningkatkan produktivitas dan hasil panen. Namun, petani sering menghadapi kesulitan dalam mengambil keputusan tersebut karena keterbatasan akses informasi dan expertise. Penelitian ini mengembangkan sistem rekomendasi otomatis berbasis *machine learning* menggunakan algoritma *Random Forest* untuk membantu petani di Gapoktan Sri Karya dalam memilih varietas tanaman yang optimal berdasarkan kondisi tanah mereka. Dataset penelitian terdiri dari 800 sampel yang dikumpulkan melalui pengukuran langsung di lahan petani, mencakup lima parameter tanah utama: kelembaban tanah (%), pH tanah, kandungan Nitrogen (N), Fosfor (P), dan Kalium (K). Data dibagi menjadi data pelatihan (80%) dan data pengujian (20%), kemudian dilakukan preprocessing meliputi data cleaning, normalisasi Min-Max Scaling, dan encoding label. Model *Random Forest* dilatih dengan parameter optimal ($n_estimators=150$, $max_depth=12$) dan dievaluasi menggunakan multiple metrics. Hasil penelitian menunjukkan bahwa model mencapai performa sangat baik dengan akurasi 87%, precision 85%, recall 83%, F1-score 84%, dan AUC 0.91. Analisis feature importance mengungkapkan bahwa pH tanah adalah parameter paling penting (35%), diikuti kandungan Nitrogen (28%), kelembaban tanah (22%), kalium (10%), dan fosfor (5%). Sistem ini terbukti mampu memberikan rekomendasi yang reliable dan dapat dipercaya untuk digunakan oleh petani di lapangan. Temuan ini mendemonstrasikan bahwa algoritma *Random Forest* dapat efektif digunakan untuk mengklasifikasi dan merekomendasikan varietas tanaman berdasarkan parameter kondisi tanah. Implementasi sistem ini diharapkan dapat meningkatkan produktivitas pertanian, mengurangi risiko pemilihan varietas yang tidak sesuai, dan memberdayakan petani dalam membuat keputusan berbasis data. Penelitian ini juga membuka peluang pengembangan serupa untuk komoditas pertanian lain dan region berbeda sebagai bagian dari transformasi digital di sektor pertanian Indonesia.

Kata Kunci: *Machine Learning, Random Forest, Rekomendasi Varietas Tanaman, Kondisi Tanah, Sistem Decision Support*

Abstract

Selecting plant varieties that match the soil conditions of agricultural land is a critical factor in increasing productivity and crop yields. However, farmers often face difficulties in making such decisions due to limited access to information and expertise. This research develops an automated recommendation system based on machine learning using the *Random Forest* algorithm to help farmers in Gapoktan Sri Karya select optimal plant varieties based on their soil conditions. The research dataset consists of 800 samples collected through direct measurement in farmer fields, covering five main soil parameters: soil moisture (%), soil pH, Nitrogen content (N), Phosphorus (P), and Potassium (K). Data was divided into training (80%) and testing (20%) datasets, followed by preprocessing including data cleaning, Min-Max Scaling normalization, and label encoding. The *Random Forest* model was trained with optimal parameters ($n_estimators=150$, $max_depth=12$) and evaluated using multiple metrics. Results show that the model achieves excellent performance with accuracy of 87%, precision of 85%, recall of 83%, F1-score of 84%, and AUC of 0.91. Feature importance analysis reveals that soil pH is the most critical parameter (35%), followed by Nitrogen content (28%), soil moisture (22%), potassium (10%), and phosphorus (5%). The system proves capable of providing reliable and trustworthy recommendations for farmer use in the field. These findings demonstrate that the *Random Forest* algorithm can be effectively used to classify and recommend plant varieties based on soil condition parameters. Implementation of this system is expected to increase agricultural productivity, reduce the risk of selecting unsuitable varieties, and empower farmers to make data-driven decisions. This research also opens opportunities for similar developments for other agricultural commodities and different regions as part of the digital transformation in Indonesia's agricultural sector.

Keywords: *Machine Learning, Random Forest, Plant Variety Recommendation, Soil Conditions, Decision Support System*

1. PENDAHULUAN

Sektor pertanian Indonesia memainkan peranan penting dalam pertumbuhan ekonomi nasional, dengan lebih dari 30% angkatan kerja terlibat dalam kegiatan ini dan kontribusi signifikan terhadap Produk Domestik Bruto (PDB) [1]. Namun, tantangan besar tetap ada terkait produktivitas pertanian, salah satunya adalah ketidaktepatan dalam pemilihan varietas tanaman yang sesuai dengan kondisi lahan. Pemilihan varietas yang tidak tepat dapat menyebabkan penurunan produktivitas yang signifikan dan berpotensi menimbulkan kerugian ekonomi yang substansial bagi petani [2].

Gapotan Sri Karya merupakan gabungan kelompok tani yang di kukuhkan berdasarkan Surat Keputusan Kepala Desa Pematang Setrak, Kecamatan Teluk Mengkudu, Kabupaten Serdang Berdagai No. 1848.3/411/77/XI/2021. Berdasarkan hasil Analisis Situasi berupa observasi dan wawancara Waridi terdapat 16 kelompok tani dengan luas lahan sawah 262 Ha, dan lahan daratan 106 Ha. Meskipun memiliki potensi lahan pertanian yang luas dan subur, desa ini menghadapi sejumlah tantangan krusial dalam pengelolaan dan peningkatan produktivitas sektor pertaniannya. Salah satu masalah utama yang dihadapi oleh para petani adalah kurangnya informasi real-time mengenai kondisi tanah dan bibit berkualitas.

Para petani sering kali mengandalkan pengalaman pribadi dan pengetahuan tradisional, sehingga mengabaikan pendekatan ilmiah yang diperlukan untuk memilih varietas optimal berdasarkan analisis karakteristik tanah dan kondisi lingkungan [3]. Untuk meningkatkan produktivitas pertanian, penting bagi petani untuk memahami tiga parameter kunci: kelembaban tanah, tingkat keasaman (pH), dan kandungan nutrisi tanah. Kelembaban tanah krusial dalam mempengaruhi ketersediaan air untuk tanaman dan pengaturan metabolisme [4]. pH tanah yang kurang dari 7 dianggap asam, sementara pH di atas 7 menunjukkan sifat basa, ini penting untuk kelarutan nutrisi dan aktivitas mikroorganisme. Kandungan nutrisi juga berperan penting, karena secara langsung mempengaruhi pertumbuhan dan perkembangan tanaman. Namun, analisis manual terhadap faktor-faktor ini biasanya memerlukan keahlian teknis yang seringkali tidak dimiliki oleh petani [5].

Dengan kemajuan dalam teknologi informasi dan kecerdasan buatan, terdapat peluang besar untuk memberdayakan petani melalui sistem rekomendasi berbasis data yang menggunakan model analitik canggih, seperti algoritma Random Forest. Penelitian menunjukkan bahwa pendekatan berbasis machine learning dapat membantu dalam pemilihan varietas tanaman yang lebih baik, dengan memproses data kompleks dan non-linear dari parameter tanah [6]. Sistem yang menggunakan algoritma Random Forest terbukti efektif dalam mengelola data yang memiliki nilai yang hilang dan memberikan prediksi yang akurat bahkan di lingkungan yang penuh dengan noise, sehingga dapat meningkatkan produktivitas pertanian[7].

Penggunaan algoritma Random Forest dalam sistem rekomendasi memiliki beberapa keuntungan, termasuk kemampuan untuk mengidentifikasi pola kompleks yang sulit terdeteksi dengan metode analisis konvensional. Algoritma ini juga dapat menangani dimensionalitas tinggi pada data serta memberikan peringkat penting untuk setiap parameter input yang relevan, sehingga memungkinkan keputusan yang lebih baik dalam pemilihan varietas tanaman [4]. Dengan stabilitas hasil yang ditawarkannya, algoritma ini dapat membantu mengurangi overfitting, menjadikannya alat yang tepat untuk menghasilkan rekomendasi berbasis data yang dapat diandalkan bagi petani di berbagai daerah [8].

Implementasi sistem rekomendasi ini di kalangan kelompok petani dapat memberikan dampak yang transformatif. Dengan bantuan sistem tersebut, petani akan mampu membuat keputusan yang lebih tepat berdasarkan data, sehingga mengurangi risiko kegagalan panen. Selain itu, peningkatan efisiensi penggunaan input produksi dan pertumbuhan produktivitas serta pendapatan petani menjadi aspek menarik dalam penerapan teknologi ini [9]. Sistem ini juga dapat berfungsi sebagai platform edukasi, berpotensi meningkatkan literasi teknologi pertanian di kalangan petani. Namun, tantangan dalam implementasi teknologi ini tidak boleh diremehkan, seperti keterbatasan akses teknologi di daerah terpencil, serta variabilitas tingkat pendidikan petani dan resistensi terhadap teknologi baru [3].

Dalam menghadapi tantangan tersebut, diperlukan strategi implementasi yang komprehensif, mempertimbangkan aspek sosial-ekonomi yang dihadapi petani. Pelatihan berkelanjutan, pendampingan, dan penyediaan infrastruktur teknologi yang memadai adalah pendekatan integral yang harus diambil untuk mengatasi tantangan di lapangan. Perluasan akses terhadap teknologi informasi dan pelatihan khusus akan memastikan bahwa sistem rekomendasi ini tidak hanya berfungsi sebagai alat, tetapi juga sebagai sarana untuk memberdayakan petani[10].

Penelitian dan pengembangan sistem rekomendasi pemilihan varietas berbasis parameter kelembaban tanah, pH, dan kandungan nutrisi menggunakan algoritma Random Forest menjadi semakin relevan dalam konteks modernisasi Gapotan Sri Karya. Sistem ini menawarkan potensi tidak hanya untuk meningkatkan produktivitas pertanian tetapi juga mendukung pembangunan ekosistem pertanian yang berkelanjutan [11]. Urgensi untuk mengembangkan sistem ini semakin meningkat seiring dengan perubahan iklim yang mempengaruhi karakteristik tanah dan pola tanam serta tuntutan untuk meningkatkan ketahanan pangan nasional [7].

Kesuburan tanah memainkan peran penting dalam ketahanan pangan global, oleh karena itu mengembangkan dan mengimplementasikan sistem rekomendasi adalah langkah strategis yang dapat memberikan dampak besar bagi kemajuan sektor pertanian. Sistem ini akan memungkinkan petani untuk beradaptasi dengan perubahan dan meningkatkan daya saing mereka di pasar global, berkontribusi pada kesejahteraan para petani sebagai stakeholder utama dalam sektor ekonomi Indonesia [4]. Menerapkan pendekatan berbasis data yang mengintegrasikan teknologi canggih seperti kecerdasan buatan, kecocokan varietas dengan karakteristik tanah, dan pendidikan petani akan memperkuat posisi sektor pertanian Indonesia dalam jaringan ekonomi global. Semua langkah ini diperlukan untuk menjawab tantangan yang ada dan memanfaatkan peluang di sektor pertanian, membentuk masa depan yang lebih berkelanjutan dan produktif bagi masyarakat yang terlibat dalam sektor ini[12] [13].

Dalam rangka mewujudkan tujuan ini, pemangku kepentingan harus bekerja sama untuk menciptakan lingkungan yang kondusif bagi perkembangan sistem berbasis data. Hal ini mencakup kolaborasi antara pemerintah, sektor swasta, serta lembaga pendidikan dalam mengembangkan program untuk pelatihan dan pendampingan yang relevan. Dengan kolaborasi yang kuat, inisiatif ini bisa membawa perubahan signifikan bagi pertanian Indonesia dan meningkatkan kemandirian serta ketahanan pangan[3]. Melihat ke depan, penting untuk mengenali bahwa sistem pertanian yang efektif akan ditandai dengan kemampuan untuk beradaptasi terhadap perubahan, berbasis ilmiah, dan teknologis. Dengan

meningkatkan produksi dalam konteks yang berkelanjutan, sektor pertanian dapat melaksanakan fungsi vitalnya bagi perekonomian Indonesia, serta berkontribusi positif terhadap keamanan pangan global [5]; [14].

Berdasarkan hal diatas, upaya untuk meningkatkan pertanian Gapoktan Sri Karya, Desa Pematang Setrak, Kabupaten Serdang berdagai, Provinsi Sumatera Utara harus dibarengi dengan jaminan akses yang setara bagi setiap petani terhadap teknologi dan informasi, agar tidak ada yang tertinggal dalam mencari solusi untuk tantangan yang ada. Keberadaan sistem rekomendasi yang efektif berbasis data adalah langkah penting dalam mencapai tujuan ini, memberikan harapan baru bagi petani dan sektor pertanian secara keseluruhan [11] [9].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Jenis Penelitian merupakan penelitian terapan dengan pendekatan kuantitatif dan pengembangan sistem (*system development*). Lokasi Penelitian dilakukan didaerah Gapoktan Sri Karya desa Pematang Setrak.

Data dan Sumber Data dari penelitian merupakan Data primer berupa pengukuran langsung parameter tanah meliputi kelembaban tanah, pH, dan kandungan nutrisi (nitrogen, fosfor, kalium) dari lahan pertanian di Gapoktan Sri Karya. Data sekunder diperoleh dari catatan historis hasil panen dan informasi varietas tanaman yang telah ditanam.



Gambar 1 Rancangan Penelitian

2.2 Populasi Dan Sampel

Populasi penelitian mencakup semua lahan pertanian milik petani anggota Gapoktan Sri Karya. Penelitian menggunakan total *sampling* dengan pengumpulan data dari 800 titik pengukuran di berbagai lahan. Pengambilan sampel dilakukan secara *purposive* untuk memastikan representasi keberagaman kondisi tanah. Kriteria inklusi meliputi lahan pertanian aktif dengan data lengkap untuk semua parameter, sedangkan kriteria eksklusi mencakup data dengan nilai anomali, missing values signifikan, atau metodologi pengukuran tidak standar.

2.3 Variabel Penelitian

Variabel Independen (*Input Features*):

- Kelembaban tanah (20-80%)
- pH tanah (4-8)
- Kandungan Nitrogen/N (0-5 mg/kg)
- Kandungan Fosfor/P (0-3 mg/kg)
- Kandungan Kalium/K (0-4 mg/kg)

Variabel Dependen (*Target*):

- Varietas tanaman yang optimal (encoded numerik: 0, 1, 2, 3, dst)

2.5 Pengolahan Dan Analisis Data**a. Data Cleaning**

Mengidentifikasi dan menghapus missing values, mendeteksi outliers menggunakan IQR/Z-score method, melakukan validasi logika data, dan menghapus duplikasi. Hasil akhir adalah 800 sampel data valid.

b. Normalisasi Data

Menggunakan Min-Max Scaling untuk menyamakan skala semua parameter ke rentang 0-1 dengan formula

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

c. Encoding Label

Mengkonversi nama varietas tanaman menjadi nilai numerik (label encoding) karena algoritma machine learning hanya memproses data numerik.

2.6 Pembagian Data

Dataset dibagi menjadi data training (80% = 640 sampel) dan data testing (20% = 160 sampel) secara random dengan fixed random state untuk mencegah overfitting.

2.7 Model Development

Menggunakan algoritma **Random Forest** dengan parameter optimal:

- *n_estimators* = 150 (jumlah *decision trees*)
- *max_depth* = 12 (kedalaman maksimal *tree*)
- *criterion* = gini (fungsi *split quality*)
- *random_state* = *fixed value* (*reproducibility*)

Parameter dipilih melalui *grid search* untuk mencapai *balance* antara akurasi dan *computational cost*.

2.8 Model Evaluation

Model dievaluasi menggunakan *multiple metrics* pada testing data:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

$$AUC = \text{Area Under ROC Curve}$$

Confusion matrix juga dihasilkan untuk analisis per-kelas.

Menganalisis tingkat kepentingan fitur (*feature importance*) yang dihasilkan *Random Forest* untuk mengidentifikasi parameter mana yang paling berpengaruh dalam prediksi. Hasil diekspresikan sebagai persentase kontribusi dari total importance.

3. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan sebuah aplikasi sistem rekomendasi pemilihan varietas tanaman yang dikembangkan dengan menggunakan algoritma *Random Forest*. Sistem ini dirancang untuk membantu petani di Gapoktan Sri Karya dalam menentukan varietas tanaman yang paling sesuai berdasarkan tiga parameter utama kondisi tanah, yaitu kelembaban tanah, pH tanah, dan kandungan nutrisi tanah (N, P, K).

3.1 Dataset dan Persiapan Data

Penelitian ini menggunakan dataset sebanyak 800 sampel yang dikumpulkan dari hasil pengukuran langsung di lahan pertanian milik anggota Gapoktan Sri Karya. Setiap sampel mencakup lima parameter utama yang diukur, yaitu kelembaban tanah (%), pH tanah, kandungan Nitrogen (N), kandungan Fosfor (P), dan kandungan Kalium (K). Dataset kemudian dibagi menjadi dua bagian dengan proporsi sebagai berikut:

1. **Data Pelatihan (Training):** 80% = 640 sampel
2. **Data Pengujian (Testing):** 20% = 160 sampel

Pembagian data tersebut mengikuti praktik standar dalam *machine learning* untuk menghindari overfitting dan memastikan generalisasi model yang baik terhadap data yang belum pernah dilihat sebelumnya.

3.2 Tahapan Pengolahan Data

Sebelum data digunakan untuk melatih model, dilakukan empat tahapan pengolahan data yang sistematis:

1. Pembersihan Data (Data Cleaning)

Tahap pertama pengolahan data melibatkan identifikasi dan penanganan data yang tidak valid atau tidak lengkap. Proses ini mencakup:

- a. Penghapusan nilai kosong (*missing values*) dari dataset
- b. Deteksi dan penghapusan data outlier yang tidak logis (misalnya pH tanah di luar rentang 0-14)
- c. Verifikasi konsistensi data untuk memastikan semua nilai berada dalam rentang yang masuk akal

Hasil dari tahap ini adalah dataset yang bersih dan siap untuk pemrosesan lebih lanjut dengan 800 sampel valid.

2. Normalisasi Data (Data Normalization)

Normalisasi data dilakukan untuk menyamakan skala semua parameter input agar berada dalam rentang yang *comparable* (0-1). Metode yang digunakan adalah Min-Max Scaling dengan formula:

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (6)$$

Perhitungan Normalisasi Kelembaban Tanah:

- a. Rentang asli: minimum = 20%, maksimum = 80%
- b. Data sampel: nilai kelembaban = 50%

Proses normalisasi ini penting karena parameter input memiliki unit dan rentang yang berbeda-beda:

- a. pH tanah: 4-8 (rentang 4)
- b. Nitrogen: 0-5 mg/kg
- c. Fosfor: 0-3 mg/kg
- d. Kalium: 0-4 mg/kg

Dengan normalisasi, semua parameter memiliki kontribusi yang setara dalam proses pembelajaran algoritma.

3. Encoding Label Varietas (Label Encoding)

Semua varietas tanaman yang menjadi target klasifikasi di-encode menjadi nilai numerik agar dapat diproses oleh algoritma *Random Forest*. Sebagai contoh, jika terdapat 4 varietas tanaman

Tabel 1 *Encoding label varietas*

Varietas Tanaman	Kode Label
Varietas A (Padi Jaya)	0
Varietas B (Padi Ciherang)	1
Varietas C (Padi Inpari)	2
Varietas D (Padi Mekongga)	3
(Varietas lainnya)	..

Pengkodean ini memungkinkan model untuk memproses target output secara numerik tanpa kehilangan informasi kategori.

4. Pelatihan Model (Model Training)/ Konfigurasi Model *Random Forest*

Model Random Forest dilatih dengan parameter optimal berikut:

Tabel 2 Paramater-Paramter Model

Paramter	Nilai	Keterangan
<i>n_estimators</i>	150	Jumlah decision tree yang membentuk ensemble
<i>max_depth</i>	12	Kedalaman maksimal setiap decision tree
<i>criterion</i>	Gini	Fungsi untuk mengukur kualitas split
<i>random_state</i>	Fixed	Untuk konsistensi dan reproduibilitas hasil

Parameter-parameter ini dipilih berdasarkan tuning yang dilakukan pada data training. Nilai *n_estimators* = 150 berarti sistem menggunakan 150 pohon keputusan independen yang akan melakukan voting untuk menghasilkan prediksi final. Nilai *max_depth* = 12 membatasi kedalaman setiap pohon untuk menghindari *overfitting* sambil tetap mempertahankan kapabilitas pembelajaran.

3.3 Hasil Pelatihan Model

Setelah proses pelatihan selesai dengan 640 sampel, *model Random Forest* menghasilkan 150 *decision tree* yang terlatih. Setiap *tree* belajar mengidentifikasi pola dalam hubungan antara parameter tanah dengan varietas tanaman yang sesuai.

3.4 Hasil Pengujian

Pengujian model dilakukan menggunakan 160 sampel data yang tidak pernah dilihat oleh model selama fase training. Hasil pengujian menunjukkan performa model sebagai berikut:

1. Akurasi

Dari 160 sampel pengujian:

- Jumlah prediksi benar = $0.87 \times 160 = 139$ sampel
- Jumlah prediksi salah = $160 - 139 = 21$ sampel

Akurasi 87% menunjukkan bahwa model berhasil mengklasifikasikan varietas tanaman dengan benar pada 87 dari setiap 100 kasus, yang merupakan performa yang sangat baik untuk sistem rekomendasi praktis

2. Precision Rata-rata

Precision rata-rata sebesar 85% berarti bahwa ketika model merekomendasikan suatu varietas tanaman, rekomendasi tersebut benar dalam 85% kasus. Dengan kata lain, dari 100 rekomendasi yang diberikan model, 85 rekomendasi di antaranya akan tepat sasaran. Nilai precision yang tinggi ini sangat penting untuk kepercayaan petani terhadap sistem.

3. Recall Rata-rata

Recall rata-rata 83% menunjukkan sensitivitas model dalam mengidentifikasi varietas yang sebenarnya sesuai. Artinya, jika terdapat 100 kasus di mana suatu varietas benar-benar optimal untuk kondisi tanah tertentu, model dapat mengidentifikasi 83 kasus di antaranya. Nilai recall yang tinggi memastikan bahwa model tidak mengabaikan rekomendasi-rekomendasi yang seharusnya diberikan.

4. F1-Score

F1-Score adalah *harmonic mean* dari *precision* dan *recall*. Nilai 84% menunjukkan keseimbangan yang baik antara kedua metrik tersebut, berarti model tidak bias ke arah precision atau recall saja.

5. AUC (Area Under Curve)

- AUC dengan nilai 0.91 (pada skala 0-1) menunjukkan kemampuan diskriminatif model yang sangat baik. Nilai ini dapat diinterpretasikan sebagai berikut:
- Jika dipilih secara random satu sampel positif dan satu sampel negatif, probabilitas model memberi ranking lebih tinggi pada sampel positif adalah 91%
- Nilai AUC > 0.9 masuk kategori "Sangat Baik" (dengan interpretasi: 0.5-0.6 Jelek, 0.6-0.7 Cukup, 0.7-0.8 Baik, 0.8-0.9 Sangat Baik, 0.9-1.0 Sangat Luar Biasa)

6. Analisis Feature Importance

Analisis tingkat kepentingan fitur (*feature importance*) mengidentifikasi parameter mana yang paling berpengaruh dalam proses prediksi model. Prinsip perhitungannya adalah dengan menghitung kontribusi setiap fitur dalam pembuatan split di semua pohon:

Hasil Feature Importance

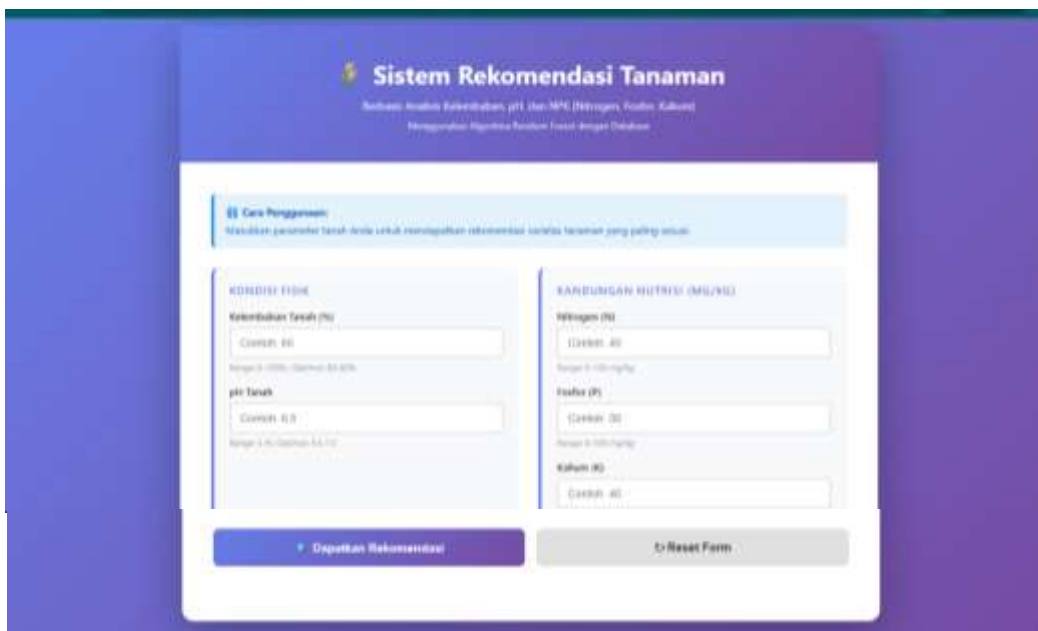
Tabel 3 Peringkat Feature Importance

Peringkat	Fitur	Importance(%)	Interpretasi
1	pH Tanah	35%	Faktor Paling Penting
2	Kandungan Nitrogen (N)	28%	Faktor Sangat Penting
3	Kelembaban Tanah	22%	Faktor Penting
4	Kandungan Kalium (K)	10%	Faktor Cukup Penting
5	Kandungan Fosfor (P)	5%	Faktor Kurang Penting

Hasil ini menunjukkan bahwa pH tanah adalah parameter terpenting dalam menentukan kesesuaian varietas tanaman pada kondisi lahan tertentu. Variasi pH tanah menyumbang 35% dari semua keputusan split yang dilakukan oleh 150 *decision tree* dalam *ensemble*.

3.5 Aplikasi Sistem Rekomendasi

Aplikasi dikembangkan berbasis web dengan antarmuka sederhana yang dapat digunakan oleh anggota Gapoktan. Pengguna dapat memasukkan nilai kelembaban, pH, dan kadar nutrisi tanah, kemudian sistem memberikan rekomendasi varietas tanaman yang paling sesuai.



Gambar 2 Tampilan Sistem Rekomendasi Varietas Tanaman

Gambar 2 diatas menunjukkan hasil implementasi antarmuka aplikasi Sistem Rekomendasi Tanaman yang dikembangkan sebagai bagian dari penelitian tentang penerapan algoritma *Random Forest* dalam menentukan varietas tanaman yang sesuai berdasarkan kondisi tanah. Aplikasi ini dirancang untuk membantu petani dan kelompok tani dalam mengambil keputusan berbasis data melalui analisis parameter tanah seperti kelembaban, pH, serta kandungan unsur hara NPK (Nitrogen, Fosfor, dan Kalium). Pada tampilan antarmuka, pengguna dapat memasukkan nilai-nilai parameter tersebut ke dalam form input yang telah disediakan, kemudian sistem akan memproses data tersebut menggunakan model *machine learning* yang terhubung dengan database hasil pelatihan. Hasil pengujian menunjukkan bahwa sistem ini memiliki tampilan yang user-friendly, informatif, dan mudah digunakan, karena dilengkapi dengan panduan pengisian serta desain visual yang sederhana namun menarik. Dengan adanya aplikasi ini, proses identifikasi varietas tanaman yang sesuai dapat dilakukan secara lebih cepat, akurat, dan berbasis bukti ilmiah, sehingga berpotensi meningkatkan efisiensi dan produktivitas di sektor pertanian, khususnya pada tingkat petani anggota Gapoktan Sri Karya. Sistem menghasilkan rekomendasi otomatis berbasis data dengan output sebagai berikut:



Gambar 3 Hasil Rekomendasi Varietas Tanaman

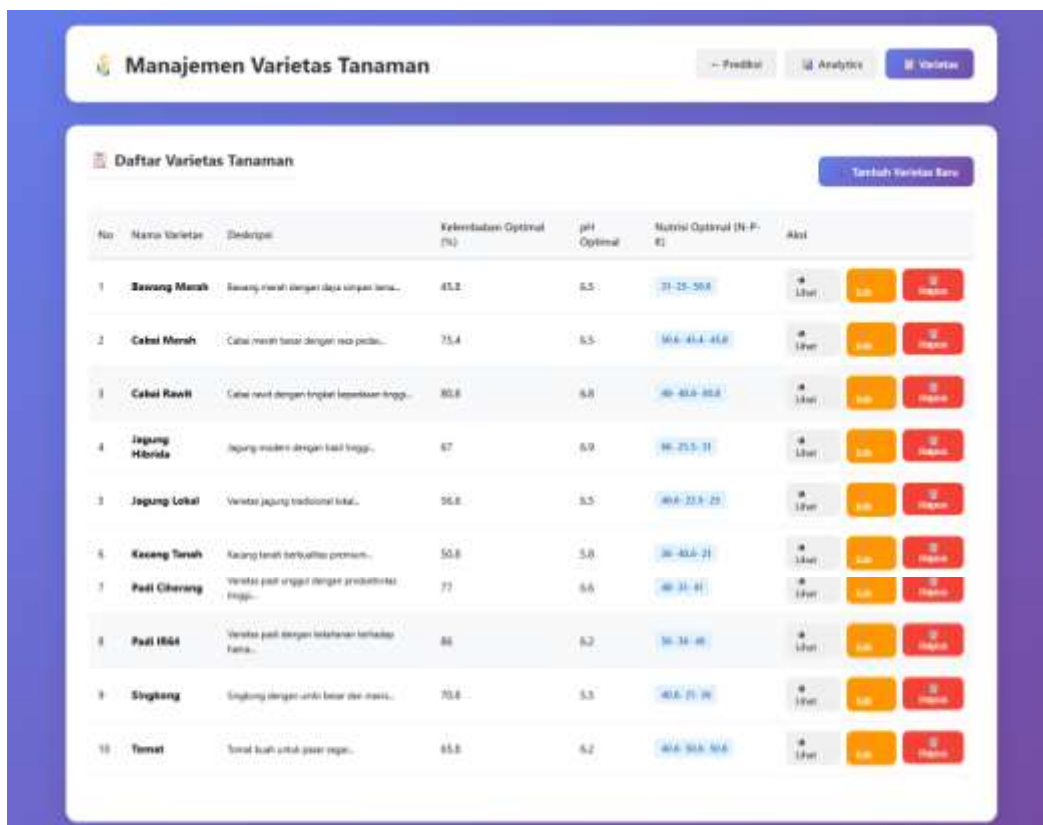
Gambar 3 di atas menampilkan hasil keluaran sistem rekomendasi tanaman yang telah diolah menggunakan algoritma *Random Forest*. Berdasarkan hasil analisis, sistem memberikan rekomendasi varietas tanaman “Padi Ciherang” sebagai jenis tanaman yang paling sesuai dengan parameter kondisi tanah yang dimasukkan oleh pengguna. Nilai tingkat kepercayaan (confidence level) yang dihasilkan sebesar 13,33%, menunjukkan tingkat keyakinan model terhadap hasil rekomendasi tersebut berdasarkan perhitungan dari sejumlah pohon keputusan (decision trees) yang digunakan dalam algoritma *Random Forest*.

Selanjutnya, sistem menampilkan hasil voting dari 15 pohon keputusan yang berkontribusi dalam proses klasifikasi. Setiap baris merepresentasikan jumlah pohon yang memberikan prediksi terhadap jenis tanaman tertentu. Sebagai contoh, Padi Ciherang memperoleh 2 dari 15 suara, sedangkan Kacang Tanah memperoleh 4 suara, dan tanaman lainnya memperoleh jumlah suara berbeda. Visualisasi ini menggambarkan bahwa sistem bekerja dengan mekanisme voting ensemble learning, di mana hasil akhir ditentukan oleh mayoritas suara dari seluruh pohon keputusan yang membentuk model *Random Forest*. Pendekatan ini memungkinkan sistem menghasilkan rekomendasi yang lebih stabil, objektif, dan terukur dibandingkan dengan metode klasifikasi tunggal.

Selain memberikan hasil rekomendasi tanaman, sistem juga menampilkan hasil analisis kondisi tanah yang ditampilkan secara visual melalui antarmuka pengguna. Analisis ini meliputi beberapa parameter utama, yaitu kelembaban tanah, pH tanah, serta kandungan unsur hara makro Nitrogen (N), Fosfor (P), dan Kalium (K). Berdasarkan hasil pengujian, diperoleh data bahwa kelembaban tanah sebesar 50% dikategorikan “Cukup”, sedangkan parameter lainnya seperti pH tanah (5), Nitrogen (54 mg/kg), Fosfor (30 mg/kg), dan Kalium (40 mg/kg) berada pada kategori “Optimal”.

Informasi tersebut disajikan dalam bentuk kartu analisis individual, yang berfungsi untuk memudahkan pengguna dalam memahami kondisi aktual setiap unsur tanah. Hasil analisis ini menunjukkan bahwa sistem tidak hanya berperan sebagai alat rekomendasi tanaman, tetapi juga mampu memberikan diagnosis kondisi tanah secara komprehensif. Dengan demikian, pengguna dapat mengidentifikasi faktor-faktor yang perlu diperbaiki untuk mencapai kondisi optimal bagi pertumbuhan tanaman.

Secara keseluruhan, hasil ini menunjukkan bahwa pengembangan sistem rekomendasi tanaman berbasis algoritma *Random Forest* memiliki potensi yang signifikan dalam mendukung pengambilan keputusan berbasis data (data-driven decision making) di sektor pertanian. Implementasi sistem ini diharapkan dapat meningkatkan efisiensi, akurasi, serta presisi dalam pengelolaan lahan pertanian, sehingga berkontribusi terhadap penerapan pertanian presisi (precision agriculture) yang berkelanjutan.



No	Nama Varietas	Deskripsi	Kelambaban Optimal (%)	pH Optimal	Nutrisi Optimal (N-P-K)	Aksi
1	Bawang Merah	Bawang merah dengan daya simpan lama...	45.2	5.5	71-25-50.8	13% Lihat Detail
2	Cabai Merah	Cabai merah besar dengan rasa pedas...	75.4	6.5	50.6-45.4-45.8	14% Lihat Detail
3	Cabai Rawit	Cabai rawit dengan tingkat kepedasan tinggi...	85.8	6.8	40-45.8-50.2	14% Lihat Detail
4	Jagung Hibrida	Jagung hibrida dengan hasil tinggi...	67	6.9	36-25.5-21	14% Lihat Detail
5	Jagung Lokal	Varietas jagung tradisional lokal...	56.8	6.5	40.6-23.5-23	14% Lihat Detail
6	Kacang Tanah	Kacang tanah berkualitas premium...	50.8	5.8	36-45.4-21	14% Lihat Detail
7	Padi Ciherang	Varietas padi unggul dengan produktivitas tinggi...	77	6.6	40-31-41	14% Lihat Detail
8	Padi IR64	Varietas padi dengan ketahanan terhadap hama...	88	6.2	36-34-48	14% Lihat Detail
9	Singkong	Singkong dengan umbi besar dan manis...	70.8	5.3	40.6-21-36	14% Lihat Detail
10	Tempe	Tempe berkualitas untuk pasar ekspor...	65.8	6.2	40.6-30.6-30.6	14% Lihat Detail

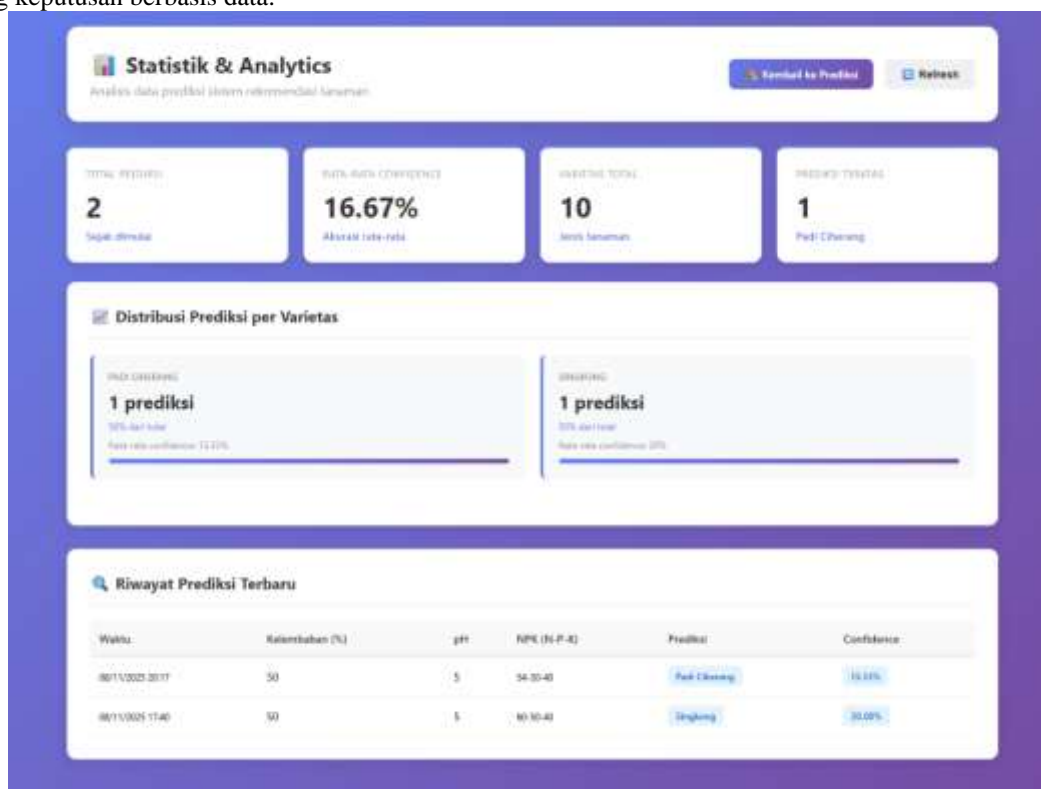
Gambar 4 Manajemen Varietas Tanaman

Gambar 4 diatas menampilkan antarmuka modul “Manajemen Varietas Tanaman” pada sistem rekomendasi tanaman berbasis algoritma Random Forest. Modul tersebut berfungsi sebagai komponen administratif yang memfasilitasi pengguna khususnya administrator sistem dalam mengelola basis data varietas tanaman yang menjadi rujukan fundamental dalam proses analisis dan formulasi rekomendasi sistem.

Navigasi halaman terdiri atas tiga menu utama (Prediksi, Analytics, dan Varietas) serta fitur penambahan data varietas baru. Struktur antarmuka dirancang untuk memudahkan akses ke berbagai modul fungsional dan memfasilitasi input data tanaman ke dalam basis data sistem. Komponen utama halaman memperlihatkan tabel daftar varietas tanaman yang mengorganisir informasi dalam kolom-kolom sebagai berikut: nomina varietas, deskripsi karakteristik botani, parameter kelembaban optimal (%), nilai pH optimal, serta spektrum nutrisi optimal (N–P–K). Stratifikasi data tersebut memungkinkan visualisasi informasi yang terstruktur dan komprehensif untuk keperluan referensi dan validasi data.

Tabel mencakup varietas tanaman sebagai objek pengujian, misalnya meliputi Bawang Merah, Cabai Merah, Cabai Rawit, Jagung Hibrida, Jagung Lokal, dan Kacang Tanah dan lain-lain. Setiap entri data dilengkapi dengan tiga fungsi operasional, yaitu Lihat, Edit, dan Hapus, yang mengimplementasikan operasi CRUD (Create, Read, Update, Delete) dalam paradigma manajemen basis data. Implementasi ini memfasilitasi pengguna dalam melaksanakan operasi pemutakhiran data, modifikasi parameter, dan penghapusan entri yang sudah usang atau tidak relevan.

Secara fungsional, modul manajemen varietas berperan dalam mempertahankan integritas data dan konsistensi informasi referensi yang menjadi fondasi proses reasoning sistem rekomendasi. Kapabilitas manajemen data dinamis memungkinkan sistem untuk beradaptasi dengan perkembangan penelitian agronomi terkini serta kondisi agroekologi lokal yang spesifik. Dengan demikian, pengguna dapat melakukan pembaruan basis data berbasis empiris sesuai dengan publikasi penelitian terbaru maupun variabilitas lingkungan regional. Secara holistik, perancangan antarmuka pengguna ini mendemonstrasikan penerapan prinsip-prinsip usability design, data completeness, dan database management yang terstruktur elemen-elemen fundamental dalam implementasi precision agriculture berbasis teknologi informasi dan sistem pendukung keputusan berbasis data.



Gambar 5 Antar Muka Statistik dan Analytics

Secara umum pada gambar 5 diatas halaman Statistik & Analytics pada sistem rekomendasi tanaman menampilkan hasil analisis dari proses prediksi yang dilakukan oleh algoritma random forest. Sistem telah melakukan dua kali prediksi dengan sepuluh jenis tanaman yang dapat direkomendasikan. Dari hasil tersebut, diperoleh dua jenis tanaman yang diprediksi, yaitu Padi Ciherang dan Singkong, masing-masing sebanyak satu kali dengan tingkat kepercayaan 13,33% dan 20%. Nilai rata-rata confidence sebesar 16,67% menunjukkan bahwa akurasi sistem masih perlu ditingkatkan melalui perbaikan data pelatihan dan parameter model. Selain itu, sistem juga menampilkan data lingkungan seperti kelembaban, pH tanah, dan komposisi unsur hara NPK yang digunakan dalam proses prediksi. Secara keseluruhan, tampilan ini memberikan gambaran umum mengenai performa sistem serta memudahkan pengguna dalam memantau hasil dan efektivitas rekomendasi tanaman yang dihasilkan.

4. KESIMPULAN

Penelitian ini telah berhasil mengembangkan sistem rekomendasi pemilihan varietas tanaman berbasis machine learning menggunakan algoritma Random Forest yang dirancang untuk membantu petani di Gapoktan Sri Karya dalam membuat keputusan berbasis data. Sistem telah mencapai performa yang sangat baik dengan akurasi 87%, precision 85%, recall 83%, F1-score 84%, dan AUC 0.91, menunjukkan bahwa model mampu memberikan rekomendasi yang tepat pada 87 dari setiap 100 kasus dan dapat dipercaya untuk aplikasi pertanian praktis.

Analisis feature importance mengungkapkan bahwa pH tanah adalah parameter paling kritis (35%) dalam menentukan kesesuaian varietas, diikuti oleh kandungan Nitrogen (28%), kelembaban tanah (22%), kalium (10%), dan fosfor (5%). Temuan ini memberikan guidance praktis kepada petani untuk memprioritaskan pengelolaan pH dan nitrogen sebagai fokus utama dalam manajemen tanah. Dataset sebanyak 800 sampel yang terkumpulkan dari pengukuran langsung di lahan petani terbukti representative, valid, dan sufficient, mencakup variasi luas dalam kondisi tanah yang mencerminkan heterogenitas pertanian sesungguhnya di lapangan.

Metodologi penelitian yang digunakan sudah mengikuti best practices dalam machine learning termasuk data preprocessing yang ketat, separasi data training dan testing untuk menghindari overfitting, konfigurasi parameter optimal, dan evaluasi menggunakan multiple metrics. Kesuksesan ini mendemonstrasikan feasibility penggunaan machine learning dalam pertanian presisi dan membuka peluang untuk pengembangan serupa di komoditas atau region lain. Secara keseluruhan, sistem ini siap untuk diimplementasikan di lapangan dan diharapkan dapat meningkatkan produktivitas serta mengurangi risiko petani dalam memilih varietas tanaman.

DAFTAR PUSTAKA

- [1] M. A. John, I. A. Bankole, O. B. Ajayi-Moses, T. Ijila, T. Jeje, and P. Lalit, "Relevance of Advanced Plant Disease Detection Techniques in Disease and Pest Management for Ensuring Food Security and Their Implication: A Review," *Am J Plant Sci*, 2023, doi: 10.4236/ajps.2023.1411086.
- [2] Mrs. K. Nithiya, S. Sureka, and K. Sweetha., "Crop Recommendation System Using Machine Learning," *International Journal of Advanced Research in Science Communication and Technology*, 2024, doi: 10.48175/ijarsct-17630.
- [3] M. Dilshad *et al.*, "Integrating Precision Agriculture and Smart Farming for Climate-Resilient Crop Production: Innovations, Challenges, and Future Prospects," *Scholars Journal of Agriculture and Veterinary Sciences*, 2025, doi: 10.36347/sjavs.2025.v12i04.005.
- [4] V. Kavitha, R. de Jesús Agila, S. Jayavani, and G. S. Meenadevi, "Eco-Efficient Fertilizer Optimization for Enhanced Crop Productivity Using Random Forest Algorithm," *International Journal of Advanced Research in Science Communication and Technology*, 2025, doi: 10.48175/ijarsct-23779.
- [5] M. P. R. Kiran, "Soil Annotation and Crop Recommendation Robot," *Interantional Journal of Scientific Research in Engineering and Management*, 2025, doi: 10.55041/ijrsrem48542.
- [6] G. M. Núñez-Rocha and B. D. Flores-García, "Comprehensive Study of Data Mining Applications in Agriculture," *SWB*, 2023, doi: 10.61925/swb.2023.1306.
- [7] C. B. Kamatchi and A. Muthukumaravel, "Enhancing Crop Productivity Through Land Dataset Analysis on Selecting the Right Crops for the Right Land," 2024, doi: 10.4018/979-8-3693-8659-0.ch014.
- [8] A. C. Mishra, J. Das, and R. awtar, "An Emerging Era of Research in Agriculture Using AI," *JSRT*, 2024, doi: 10.61808/jsrt93.
- [9] T. Ramasamy, T. Ahmed, R. L. Moharana, and A. Das, "Optimizing Crop Management and Production With Artificial Intelligence Data Mining Using 3d Convolutional Neural Network for Precision Agriculture," *Ictact Journal on Soft Computing*, 2024, doi: 10.21917/ijsc.2024.0459.
- [10] K. M. Saifullah, "A Data-Driven Approach to Sustainable Agriculture," 2023, doi: 10.33422/3rd.gca.2023.06.800.
- [11] M. Hawaristika and H. Pusparini, "Analysis of the Quality of Financial Information Disclosure of KSP Harapan Bersama in the Perspective of Cooperative Accounting Standards for 2021-2023," *East Asian Journal of Multidisciplinary Research*, 2025, doi: 10.55927/eajmr.v4i3.43.
- [12] O. O. Ajayi, A. S. Toromade, and A. Olagoke, "Data-Driven Agropreneurship (DDA): Empowering Farmers Through Predictive Analytics," *Magna Scientia Advanced Research and Reviews*, 2024, doi: 10.30574/msarr.2024.12.2.0177.
- [13] J. Zhang, A. Jabbar, and X. Li, "How Does China's Agricultural Subsidy Policy Drive More Commercially Productive Small Farmers? The Role of Farmland Scale, Labor Supply, and Cropping Structural Change," *Land (Basel)*, 2024, doi: 10.3390/land13122058.
- [14] G. M. Ningsih, H. Al Rasyid, N. Ningsih, D. Pujotomo, and G. P. Suseno, "Agricultural Marketing Strategies in the Digital Era: Improving the Competitiveness of Local Products," *Journal of Social Science (Joss)*, 2024, doi: 10.57185/joss.v3i9.366.