

ANALISIS PERBANDINGAN RANDOM FOREST DAN SUPPORT VECTOR MACHINE (SVM) DALAM SISTEM REKOMENDASI JENIS TANAMAN

COMPARATIVE ANALYSIS OF RANDOM FOREST AND SUPPORT VECTOR MACHINE (SVM) IN PLANT TYPE RECOMMENDATION SYSTEM

Anis'syah Mahdarani¹, Irma Handayani²

Universitas Teknologi Yogyakarta^{1,2}

anissyahmahdarani@gmail.com¹

ABSTRACT

Selecting suitable crop types based on environmental and soil nutrient conditions is an important factor in improving agricultural productivity. This study aims to analyze and compare the performance of two Machine Learning algorithms, namely Random Forest and Support Vector Machine (SVM), in a crop recommendation system based on environmental and soil nutrient factors. The dataset includes parameters such as temperature, humidity, wind speed, soil pH, and nutrient contents of nitrogen, phosphorus, and potassium. The evaluation results show that the Random Forest algorithm achieved the highest accuracy of 89.03%, slightly outperforming SVM with 88.51%. The feature importance analysis indicates that Nitrogen (N), Potassium (K), and Phosphorus (P) are the most influential factors. Therefore, the Random Forest algorithm is considered more optimal for implementation in a crop recommendation system based on environmental and soil nutrient data.

Keywords: *Random Forest, SVM, Crop Recommendation, Machine Learning, Agriculture.*

ABSTRAK

Pemilihan jenis tanaman yang sesuai dengan kondisi lingkungan dan nutrisi tanah merupakan faktor penting dalam meningkatkan produktivitas pertanian. Penelitian ini bertujuan untuk menganalisis dan membandingkan performa dua algoritma *Machine Learning*, yaitu *Random Forest* dan *Support Vector Machine (SVM)*, dalam sistem rekomendasi jenis tanaman berdasarkan faktor lingkungan dan nutrisi tanah. Dataset yang digunakan mencakup parameter seperti suhu, kelembapan, kecepatan angin, pH tanah, serta kandungan *nitrogen*, *fosfor*, dan *kalium*. Hasil evaluasi menunjukkan bahwa algoritma *Random Forest* memiliki akurasi tertinggi sebesar 89,03%, sedikit lebih baik dibandingkan *SVM* dengan akurasi 88,51%. *Analisis feature importance* menunjukkan bahwa *Nitrogen (N)*, *Kalium (K)*, dan *Fosfor (P)* merupakan faktor paling berpengaruh. Dengan demikian, algoritma *Random Forest* dinilai lebih optimal untuk diterapkan dalam sistem rekomendasi tanaman berdasarkan data lingkungan dan nutrisi tanah.

Kata Kunci: *Random Forest, SVM, Rekomendasi Tanaman, Machine Learning, Pertanian.*

PENDAHULUAN

Pertanian merupakan sektor strategis dalam menjaga ketahanan pangan dan mendukung perekonomian nasional. Pemilihan jenis tanaman yang sesuai dengan kondisi lingkungan dan nutrisi tanah menjadi faktor penting dalam meningkatkan produktivitas serta efisiensi lahan pertanian. Menurut *Food and Agriculture Organization (FAO, 2024)*, transformasi sistem pangan global perlu dimulai dari peningkatan efisiensi produksi melalui praktik pertanian berkelanjutan yang didukung oleh teknologi digital dan analisis data. Upaya ini sejalan dengan visi FAO untuk “Mewujudkan empat pilar utama, yaitu produksi yang lebih baik,

nutrisi yang lebih baik, lingkungan yang lebih baik, dan kehidupan yang lebih baik melalui penerapan teknologi dalam pengambilan keputusan di sektor pertanian”. Direktur Jenderal FAO, Qu Dongyu, menegaskan bahwa digitalisasi dan pemanfaatan data ilmiah merupakan masa depan pertanian dunia karena dapat membantu petani mengambil keputusan yang lebih tepat berdasarkan kondisi lahan dan lingkungan mereka (Herman, 2022). Sejalan dengan hal tersebut, Kementerian Pertanian Republik Indonesia (Sekretariat Jendral, 2023) dalam *Analisis Komoditas Pangan Strategis* menekankan pentingnya pemanfaatan data dan indikator lingkungan dalam menganalisis komoditas unggulan

nasional serta mendukung perencanaan produksi yang berkelanjutan. Oleh karena itu, penerapan metode berbasis data dan *Machine Learning* menjadi langkah strategis dalam mendukung pengambilan keputusan dalam pemilihan jenis tanaman yang optimal berdasarkan faktor lingkungan dan nutrisi tanah.

Meskipun pertanian memiliki peran penting dalam menjaga ketahanan pangan, praktik di lapangan masih banyak bergantung pada pengalaman dan kebiasaan petani tanpa mempertimbangkan faktor ilmiah seperti pH tanah, unsur hara (nitrogen, fosfor, dan kalium), suhu, kelembapan, serta kecepatan angin. Pendekatan konvensional ini sering mengakibatkan ketidaksesuaian antara jenis tanaman dan kondisi lahan, sehingga berdampak pada penurunan hasil panen dan peningkatan risiko gagal panen. Penelitian yang dilakukan oleh (Occelli et al., 2021) menunjukkan bahwa pemahaman ilmiah mengenai pengelolaan tanah berpengaruh langsung terhadap kemampuan petani dalam menjaga kesuburan lahan, terutama dalam pengelolaan unsur makronutrien seperti N, P, dan K. Di Indonesia, praktik pemupukan tanpa uji tanah yang tepat masih umum dilakukan sehingga sering menimbulkan ketidaksesuaian antara jenis tanaman dan kondisi lahan (Kencana, 2025). Kondisi ini turut memperburuk krisis kesuburan tanah yang mulai menjadi ancaman serius terhadap produktivitas pertanian nasional (Anonymous, 2025). Selain itu, perubahan iklim global turut memperparah kondisi pertanian dengan memengaruhi stabilitas suhu, curah hujan, dan kelembapan tanah, yang pada akhirnya menimbulkan ketidakpastian terhadap hasil produksi pangan (Badan Pangan Nasional, 2023). Dalam konteks tersebut, diperlukan sistem pertanian berbasis data dan analisis prediktif yang mampu memberikan rekomendasi tanaman sesuai dengan karakteristik lahan dan kondisi iklim secara *real time*.

Seiring dengan perkembangan teknologi, penerapan kecerdasan buatan

(*Artificial Intelligence* - AI) dan *Machine Learning* (ML) dalam bidang pertanian telah menjadi inovasi penting dalam mendorong efisiensi dan keberlanjutan sistem pangan global. Teknologi ini memungkinkan proses pertanian dilakukan secara cerdas melalui analisis data lingkungan, kondisi tanah, serta prediksi hasil panen yang lebih akurat. Berbagai penelitian sebelumnya menunjukkan efektivitas AI dalam mendukung pertanian presisi. Menurut (Adewusi et al., 2024), integrasi AI dalam pertanian presisi mampu meningkatkan efisiensi penggunaan sumber daya seperti air dan pupuk, serta membantu petani mengambil keputusan berbasis data secara *real-time*. Sementara itu, (Senapaty et al., 2024) menemukan bahwa model ML dapat digunakan untuk merekomendasikan jenis tanaman yang sesuai dengan karakteristik tanah, tingkat keasaman (pH), dan unsur hara seperti nitrogen, fosfor, serta kalium. Penelitian lain oleh (Ali et al., 2025) juga menegaskan bahwa penerapan teknik ML dan *deep learning* berpotensi memperkuat ketahanan pangan melalui prediksi hasil panen dan pengelolaan lahan yang adaptif terhadap perubahan iklim. Dengan demikian, pengembangan sistem pertanian cerdas berbasis pembelajaran mesin menjadi langkah strategis dalam mewujudkan pertanian yang efisien, akurat, dan berkelanjutan.

Perkembangan *Machine Learning* telah membuka peluang besar dalam penerapan teknologi cerdas di bidang pertanian, khususnya untuk sistem prediksi dan rekomendasi tanaman. Berbagai algoritma seperti *Decision Tree*, *Naïve Bayes*, *K-Nearest Neighbors (KNN)*, *Random Forest*, dan *Support Vector Machine (SVM)* telah banyak digunakan untuk mengoptimalkan pengambilan keputusan berbasis data lingkungan dan nutrisi tanah. Penelitian oleh (Burla et al., 2024) menunjukkan bahwa algoritma *Random Forest* mencapai akurasi hingga 99% dalam sistem rekomendasi tanaman berbasis data lingkungan dan karakteristik

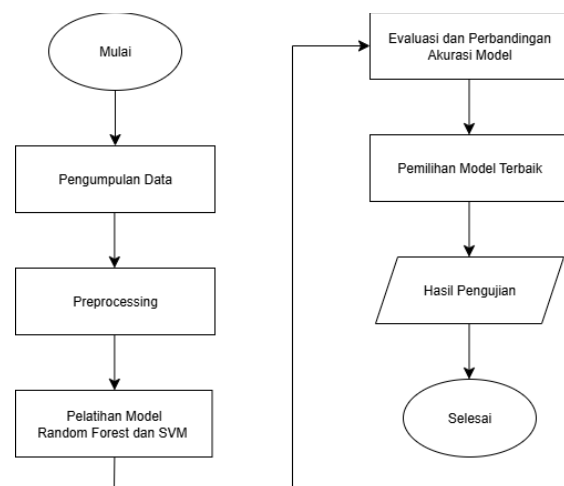
tanah, sementara (Rahman et al., 2024) menemukan bahwa algoritma ini memiliki *Mean Absolute Error (MAE)* yang lebih rendah dibandingkan *XGBoost Regressor*, sehingga lebih stabil untuk prediksi berbasis nutrisi tanah. Di sisi lain, beberapa penelitian terkini juga menyoroiti keunggulan *Support Vector Machine (SVM)* dalam menangani data non-linear di sektor pertanian. Misalnya, penelitian oleh (Siregar & Nasution, 2024) berhasil memprediksi hasil panen padi menggunakan algoritma SVM dengan akurasi tinggi pada data lingkungan di Sumatera Utara, sedangkan (Kurniawan et al., 2023) menerapkan SVM untuk prediksi hasil panen salak dengan hasil yang signifikan. Selain itu, (Yasin et al., 2025) melakukan studi komparatif antara *Support Vector Regression (SVR)*, *Random Forest*, dan *XGBoost* dalam prediksi produktivitas padi, dan menemukan bahwa SVR menunjukkan kemampuan generalisasi yang kuat pada data non-linear. Berdasarkan hasil-hasil tersebut, baik *Random Forest* maupun SVM terbukti memiliki performa tinggi di bidang pertanian. *Random Forest* unggul dalam stabilitas dan penanganan data kompleks, sedangkan SVM memiliki kemampuan generalisasi yang baik untuk data non-linear. Namun, perbandingan kinerja kedua algoritma tersebut dalam konteks sistem rekomendasi jenis tanaman berbasis faktor lingkungan dan nutrisi tanah masih jarang dilakukan. Oleh karena itu, penelitian ini bertujuan untuk menganalisis dan membandingkan performa kedua algoritma tersebut guna menghasilkan sistem rekomendasi tanaman yang lebih akurat dan efisien, serta berkontribusi terhadap pengembangan sistem pertanian cerdas di Indonesia.

Dengan demikian, penelitian ini diharapkan dapat memberikan manfaat baik secara praktis maupun akademis. Secara praktis, sistem rekomendasi yang dikembangkan dapat membantu petani dalam menentukan jenis tanaman yang sesuai dengan kondisi lahan dan iklim

secara efisien, sehingga meningkatkan produktivitas dan mengurangi risiko gagal panen. Secara akademis, hasil penelitian ini diharapkan dapat menjadi rujukan bagi pengembangan model *Machine Learning* di bidang pertanian presisi dan pengambilan keputusan berbasis data lingkungan serta nutrisi tanah.

METODE

Pada bagian ini dijelaskan tahapan-tahapan penelitian yang dilakukan, mulai dari proses pengumpulan *dataset*, *preprocessing* data, pelatihan model menggunakan algoritma *Random Forest* dan *Support Vector Machine (SVM)*, hingga tahap evaluasi dan pemilihan model terbaik. Secara umum, tahapan penelitian dalam studi ini dapat dilihat pada Gambar 1.



Gambar 1. Tahap Penelitian

Berdasarkan Gambar 1, penelitian ini dimulai dengan tahap pengumpulan data yang terdiri atas data faktor lingkungan dan nutrisi tanah. Setelah itu, dilakukan *preprocessing* data yang meliputi pembersihan data (*data cleaning*), pengkodean (*encoding*), normalisasi, serta penanganan ketidakseimbangan data menggunakan metode *SMOTE*. Tahap berikutnya adalah pelatihan model menggunakan dua algoritma, yaitu *Random Forest* dan *SVM*, yang kemudian diuji untuk membandingkan tingkat akurasi dan performanya. Hasil perbandingan tersebut digunakan untuk menentukan model

terbaik, yang selanjutnya disimpan dan digunakan dalam sistem rekomendasi jenis tanaman berbasis web.

Dataset

Dataset yang digunakan dalam penelitian ini berasal dari platform Kaggle dengan judul “*Crop Yield and Environmental Factors (2014–2023)*” yang dikembangkan oleh Madhan Kumar (2023) dan dapat diakses melalui tautan <https://www.kaggle.com/datasets/madhankumar789/crop-yield-and-environmental-factors-2014-2023>. Dataset ini berisi data pertanian global tahun 2014-2023 yang mencakup berbagai parameter lingkungan, kondisi tanah, dan jenis tanaman yang dibudidayakan. Pada penelitian ini hanya digunakan atribut yang relevan dengan faktor lingkungan dan nutrisi tanah, yaitu suhu, kelembapan, kecepatan angin, pH tanah, serta kandungan *nitrogen* (N), *fosfor* (P), *kalium* (K), dan jenis tanah, sedangkan atribut waktu diekstraksi menjadi tahun, bulan, dan hari, untuk memperkaya variasi data temporal dalam proses analisis. Atribut target yang digunakan adalah jenis tanaman (*Crop Type*), difokuskan pada sembilan tanaman yang umum dibudidayakan di Indonesia, yaitu padi, jagung, gandum, kedelai, tebu, tomat, kentang, bunga matahari, dan kapas. Untuk menyesuaikan konteks lokal, nama tanaman dalam bahasa Inggris diubah ke bahasa Indonesia menggunakan *mapping dictionary*, kemudian data dibersihkan dari nilai kosong dan entri yang tidak relevan agar dataset yang digunakan valid, representatif, dan siap digunakan dalam proses pelatihan model *Machine Learning*.

Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan data siap digunakan dalam proses pelatihan model *Machine Learning*. Pada tahap ini, dilakukan serangkaian proses untuk mempersiapkan data agar lebih terstruktur dan mudah diolah oleh algoritma. Sejalan dengan pendapat (Hamdani et al., 2024), tahap data

preprocessing berperan penting dalam mengubah data mentah menjadi bentuk yang lebih terorganisir, efisien, dan siap digunakan sebelum masuk ke tahap pengolahan atau analisis lebih lanjut. Proses ini umumnya mencakup kegiatan seperti integrasi, transformasi, pembersihan, dan reduksi data guna meningkatkan kualitas dataset yang akan diproses. Beberapa langkah yang dilakukan adalah sebagai berikut:

1. Encoding Label Tanaman
Kelas target (*Crop_Type*) dikonversi menjadi bentuk numerik menggunakan metode `LabelEncoder()` agar dapat diproses oleh algoritma *Machine Learning*.
2. Ekstraksi Fitur Waktu
Kolom *Date* diubah menjadi tiga fitur baru, yaitu *Year*, *Month*, dan *Day*, untuk menangkap pengaruh waktu terhadap kondisi lingkungan pertanian.
3. Encoding Jenis Tanah
Atribut kategorikal *Soil_Type* dikodekan menggunakan `LabelEncoder()` agar dapat diproses dalam bentuk numerik.
4. Normalisasi Data
Proses normalisasi dilakukan menggunakan `StandardScaler()` untuk menyeragamkan skala antar fitur numerik sehingga tidak ada fitur yang mendominasi perhitungan jarak, terutama pada algoritma SVM yang sensitif terhadap skala data.
5. Pembagian Dataset
Dataset dibagi menjadi *training set* (80%) dan *testing set* (20%) menggunakan `train_test_split()` dengan parameter stratify agar proporsi kelas tanaman tetap seimbang di kedua subset data.
6. Penanganan Ketidakseimbangan Data
Untuk mengatasi ketidakseimbangan jumlah data antar kelas tanaman, digunakan metode *Synthetic Minority Over-sampling Technique (SMOTE)* pada data latih guna menghasilkan sampel sintesis pada kelas minoritas.

Langkah-langkah ini memastikan dataset memiliki kualitas yang baik,

terstandardisasi, dan seimbang sehingga dapat meningkatkan performa model *Random Forest* dan *Support Vector Machine (SVM)* pada tahap pelatihan berikutnya.

Algoritma Machine Learning

Penelitian ini menggunakan dua algoritma *Machine Learning*, yaitu *Random Forest* dan *Support Vector Machine (SVM)*, untuk menganalisis serta membandingkan performa dalam sistem rekomendasi jenis tanaman berdasarkan faktor lingkungan dan nutrisi tanah. Kedua algoritma ini dipilih karena memiliki karakteristik yang saling melengkapi dalam menangani data kompleks dan non-linear, sehingga relevan untuk diterapkan pada permasalahan klasifikasi multi-kelas seperti prediksi jenis tanaman.

1. Random Forest

Random Forest adalah salah satu metode *ensemble learning* yang membangun sejumlah *decision tree* dan menggabungkan hasilnya melalui mekanisme *voting* untuk memperoleh keputusan akhir yang lebih akurat. Algoritma ini memanfaatkan pemilihan fitur dan data secara acak pada setiap pohon, sehingga dapat mengurangi risiko *overfitting* dan meningkatkan generalisasi model (Dachi & Sitompul, 2023).

Dalam penelitian ini, model dilatih menggunakan parameter `random_state=42` dan `class_weight="balanced"` untuk menjaga kestabilan hasil serta menyesuaikan bobot antar kelas tanaman yang tidak seimbang. Metode ini juga dikenal efisien dalam menangani data berdimensi tinggi, memberikan interpretasi melalui nilai *feature importance*, serta mampu memodelkan hubungan non-linear antar variabel.

2. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma pembelajaran terawasi yang berfungsi memisahkan data ke dalam dua atau lebih kelas dengan membentuk

hyperplane terbaik yang memiliki margin maksimum antar kelas (Zalvadila et al., 2023). Dalam penelitian ini digunakan fungsi kernel *Radial Basis Function (RBF)* karena mampu menangani data non-linear dengan baik. Parameter yang digunakan meliputi `kernel="rbf"`, `C=1`, `gamma="scale"`, dan `class_weight="balanced"`.

Kedua model kemudian dilatih menggunakan data hasil *SMOTE* dan dibandingkan berdasarkan metrik evaluasi yang meliputi *akurasi*, *presisi*, *recall*, dan *F1-score*. Model dengan performa terbaik dianalisis lebih lanjut menggunakan nilai *feature importance* untuk mengetahui kontribusi tiap fitur terhadap hasil rekomendasi jenis tanaman.

Evaluasi Model

Evaluasi model dilakukan untuk mengukur kinerja algoritma *Random Forest* dan *Support Vector Machine (SVM)* dalam melakukan klasifikasi jenis tanaman berdasarkan faktor lingkungan dan nutrisi tanah. Proses evaluasi ini bertujuan untuk mengetahui sejauh mana model mampu memberikan rekomendasi tanaman yang akurat dan konsisten terhadap data uji yang belum pernah dilihat sebelumnya. Empat metrik utama digunakan dalam proses evaluasi, yaitu:

1. Accuracy

Accuracy merupakan proses untuk mengukur seberapa sering model menghasilkan prediksi yang benar terhadap keseluruhan data uji. Rumus untuk menghitung nilai akurasi dapat dituliskan sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Keterangan:

- TP (True Positive) : jumlah prediksi benar pada kelas positif
- TN (True Negative) : jumlah prediksi benar pada kelas negatif
- FP (False Positive) : jumlah data negatif yang salah diprediksi positif

- FN (False Negative) : jumlah data positif yang salah diprediksi negatif
- Metrik ini memberikan gambaran umum mengenai performa keseluruhan model terhadap data uji.

2. Precision

Precision merupakan metrik yang digunakan untuk menunjukkan proporsi data yang diprediksi positif dan benar-benar positif, atau tingkat ketepatan model dalam melakukan klasifikasi. Rumusnya dituliskan sebagai berikut:

$$Precision = \frac{TP}{TP + FP}$$

Nilai *precision* yang tinggi menunjukkan bahwa model memiliki tingkat kesalahan positif palsu (*false positive*) yang rendah.

3. Recall

Recall atau sensitivitas menunjukkan kemampuan model dalam mengenali seluruh data yang benar-benar positif dari keseluruhan data aktual. Rumus *recall* adalah sebagai berikut:

$$Recall = \frac{TP}{TP + FN}$$

Nilai *recall* yang tinggi menunjukkan bahwa model mampu mendeteksi sebagian besar data positif secara benar.

4. F1-Score

F1-Score merupakan rata-rata harmonik dari nilai *precision* dan *recall* yang berfungsi untuk menilai keseimbangan kinerja model, terutama saat terdapat distribusi kelas yang tidak seimbang. Persamaannya dapat ditulis sebagai berikut:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Nilai *F1-Score* yang mendekati 1 menunjukkan bahwa model memiliki keseimbangan baik antara kemampuan mendeteksi data positif dan menghindari kesalahan prediksi.

Model dengan kombinasi nilai *akurasi*, *presisi*, *recall*, dan *F1-score*

terbaik kemudian dianalisis lebih lanjut menggunakan *Feature Importance* guna mengidentifikasi faktor lingkungan dan nutrisi tanah yang paling berpengaruh terhadap hasil prediksi. Berdasarkan hasil pengujian, algoritma *Random Forest* menunjukkan performa yang lebih stabil dan akurat dibandingkan dengan *Support Vector Machine (SVM)*, sehingga dipilih sebagai model terbaik dalam sistem rekomendasi tanaman ini (Saputra et al., 2025).

Penyimpanan Model dan Encoder

Model terbaik yang diperoleh dari hasil evaluasi kemudian disimpan dalam format file .pkl (Pickle) menggunakan pustaka *joblib*. Proses penyimpanan ini dilakukan agar model dapat dimanfaatkan kembali tanpa perlu melakukan pelatihan ulang (*retraining*). Beberapa komponen penting yang disimpan meliputi:

1. Model terbaik (*best_model.pkl*), yaitu model *Random Forest* yang menunjukkan performa akurasi tertinggi dibandingkan dengan *SVM*.
2. *scaler.pkl*, digunakan untuk menyimpan hasil standarisasi fitur numerik agar proses normalisasi pada saat prediksi selanjutnya tetap konsisten dengan tahap pelatihan.
3. *soil_encoder.pkl* dan *label_encoder.pkl*, masing-masing berfungsi untuk mengonversi data kategorikal (jenis tanah dan label tanaman) menjadi bentuk numerik yang dapat diproses oleh model.
4. *feature_order.pkl*, berisi urutan fitur yang digunakan pada tahap pelatihan, guna memastikan kesesuaian input antara model dan antarmuka aplikasi.

Penyimpanan seluruh komponen tersebut bertujuan untuk mendukung penerapan sistem rekomendasi tanaman berbasis web menggunakan *Flask*. Dengan cara ini, model yang telah dilatih dapat langsung diintegrasikan ke dalam aplikasi tanpa perlu dilakukan pelatihan ulang, sehingga mempercepat waktu respon dan menjaga konsistensi hasil prediksi.

Pendekatan ini sejalan dengan praktik umum dalam implementasi *Machine Learning Deployment*, di mana model yang telah diverifikasi akan diserialisasi untuk keperluan inferensi di lingkungan produksi.

HASIL DAN PEMBAHASAN

Hasil Evaluasi Model

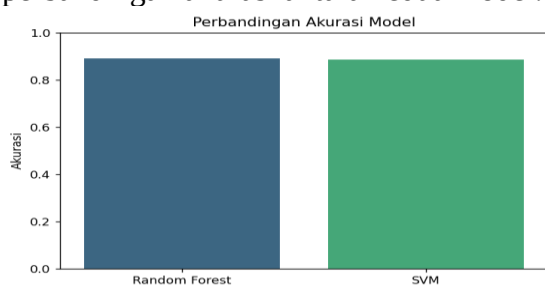
Penelitian ini membandingkan performa dua algoritma *Machine Learning*, yaitu *Random Forest* dan *Support Vector Machine (SVM)*, dalam sistem rekomendasi jenis tanaman berbasis faktor lingkungan dan nutrisi tanah. Kedua model diuji menggunakan data uji sebesar 20% dari total dataset setelah melalui tahap preprocessing dan penyeimbangan data menggunakan metode *SMOTE*.

Hasil evaluasi menunjukkan bahwa *Random Forest* memiliki akurasi sebesar 0.8903, sedikit lebih tinggi dibandingkan *SVM* yang memiliki akurasi 0.8851. Meskipun selisihnya relatif kecil, hasil ini menunjukkan bahwa *Random Forest* lebih stabil dalam menangani data dengan variasi tinggi dan hubungan non-linear antar fitur. Perbandingan hasil evaluasi kedua model disajikan pada Tabel 1.

Tabel 1. Hasil Evaluasi Model

	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1Score</i>
Random Forest	0.8903	0.89	0.89	0.89
SVM	0.8851	0.89	0.89	0.89

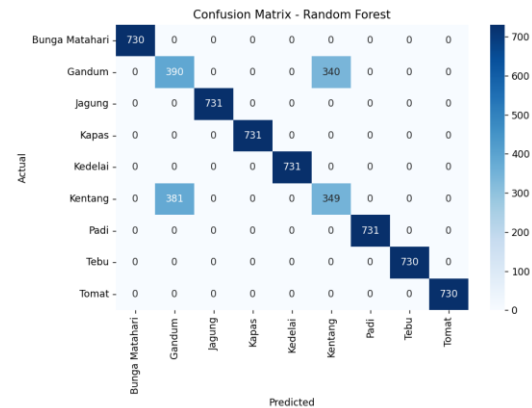
Secara umum, kedua algoritma menunjukkan performa yang baik dengan nilai *precision*, *recall*, dan *F1-score* yang relatif seimbang pada semua metrik. Namun, *Random Forest* memberikan hasil yang sedikit lebih konsisten pada kelas tanaman dengan distribusi data tidak merata. Gambar 2 menunjukkan perbandingan akurasi antara kedua model.



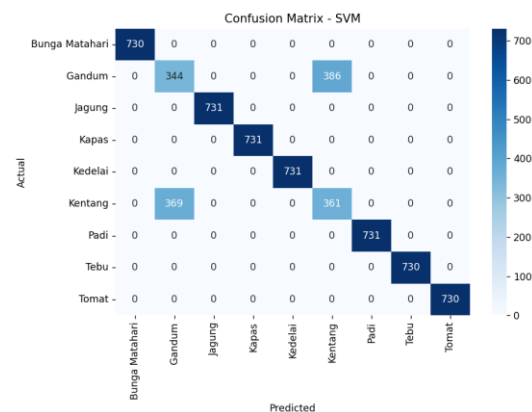
Gambar 2. Perbandingan Akurasi

Analisis Confusion Matrix

Analisis lebih lanjutan dilakukan menggunakan confusion matrix untuk mengidentifikasi tingkat keberhasilan dan pola kesalahan klasifikasi antar jenis tanaman pada kedua model yang diuji.



Gambar 3. Confusion Matrix Random Forest



Gambar 4. Confusion Matrix Support Vector Machine (SVM)

Berdasarkan Gambar 3 dan Gambar 4, terlihat bahwa sebagian besar jenis tanaman seperti Bunga Matahari, Jagung, Kapas, Kedelai, Padi, Tebu, dan Tomat berhasil diklasifikasikan dengan benar oleh kedua model. Hal ini ditunjukkan dengan nilai diagonal utama confusion matrix yang mendekati jumlah total data untuk masing-masing kelas, menunjukkan akurasi yang sangat tinggi (*precision* dan *recall* $\approx 1,00$) pada kelas-kelas tersebut.

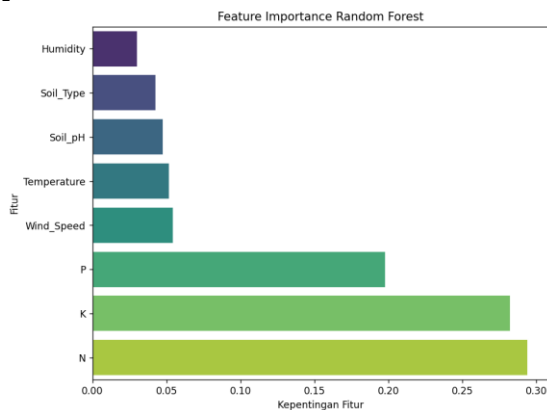
Namun, terdapat kesalahan klasifikasi (*misclassification*) yang cukup menonjol pada dua jenis tanaman, yaitu Gandum dan Kentang. Kedua model sama-sama mengalami kesulitan dalam membedakan kedua kelas ini. Pada model

Random Forest, sebagian data Gandum (sekitar 47%) salah diklasifikasikan sebagai Kentang, dan sebagian data Kentang (sekitar 48%) juga salah diprediksi sebagai Gandum. Pada model *SVM*, pola kesalahan serupa terjadi, namun dengan akurasi yang sedikit lebih rendah dibandingkan *Random Forest*.

Kemiripan pola nilai pada faktor lingkungan dan nutrisi tanah antara tanaman Gandum dan Kentang diduga menjadi penyebab utama terjadinya kesalahan klasifikasi pada kedua model. Meskipun demikian, model *Random Forest* menunjukkan kinerja yang lebih stabil dan konsisten dalam meminimalkan kesalahan klasifikasi antar kelas dibandingkan dengan model *SVM*.

Analisis Feature Importance

Setelah memperoleh hasil evaluasi dan confusion matrix, dilakukan analisis *feature importance* untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap hasil rekomendasi jenis tanaman pada model *Random Forest*.



Gambar 5. Feature Importance Random Forest

Berdasarkan hasil visualisasi pada Gambar 5, terlihat bahwa *Nitrogen* (N), *Kalium* (K), dan *Fosfor* (P) memiliki tingkat kepentingan tertinggi dalam menentukan hasil prediksi model. Ketiga fitur ini memberikan pengaruh paling dominan terhadap keputusan sistem dalam merekomendasikan jenis tanaman yang sesuai dengan kondisi lahan.

Selanjutnya, fitur, pH tanah (*Soil_pH*), dan Jenis Tanah (*Soil_Type*),

Suhu (*Temperature*), Kecepatan Angin (*Wind_Speed*), dan Kelembapan (*Humidity*) juga memberikan kontribusi terhadap hasil prediksi, meskipun dengan tingkat pengaruh yang lebih rendah dibandingkan tiga fitur utama sebelumnya.

Temuan ini mengindikasikan bahwa unsur hara makro (N, P, dan K) serta kualitas tanah merupakan komponen paling penting dalam menentukan kecocokan lahan terhadap jenis tanaman tertentu. Kandungan unsur hara yang seimbang dan kualitas tanah yang baik mendukung pertumbuhan tanaman secara optimal, sehingga berdampak langsung pada keberhasilan budidaya.

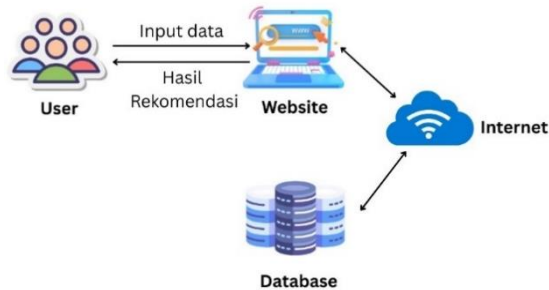
Secara keseluruhan, hasil *analisis feature importance* memperkuat bahwa data nutrisi tanah dan faktor lingkungan merupakan elemen kunci dalam sistem rekomendasi tanaman berbasis pembelajaran mesin. Model *Random Forest* tidak hanya menghasilkan prediksi dengan akurasi tinggi, tetapi juga mampu memberikan interpretasi yang jelas mengenai fitur-fitur yang paling memengaruhi keputusan model.

IMPLEMENTASI SISTEM

Implementasi sistem dilakukan untuk mewujudkan aplikasi rekomendasi jenis tanaman berbasis Machine Learning yang dapat diakses melalui web. Sistem ini dikembangkan menggunakan *framework Flask (Python)* dan diintegrasikan dengan database *MySQL* sebagai media penyimpanan data pengguna serta riwayat hasil prediksi. Model pembelajaran mesin yang digunakan berasal dari hasil pelatihan algoritma *Random Forest* dengan akurasi terbaik (89,03%), yang disimpan dalam format *.pkl* agar dapat dimuat ulang tanpa proses pelatihan ulang. Struktur sistem terdiri dari tiga komponen utama, yaitu:

1. Antarmuka pengguna (*frontend*) yang menampilkan halaman beranda, form input, hasil rekomendasi, dan riwayat prediksi.

2. Model *Machine Learning* (*backend*) yang memproses data input dan menghasilkan prediksi jenis tanaman.
3. Database *MySQL*, yang menyimpan data pengguna, input faktor lingkungan dan nutrisi tanah, serta hasil rekomendasi tanaman.



Gambar 6. Arsitektur Sistem Rekomendasi

Gambar 6 menunjukkan alur komunikasi antara pengguna (*user*), *website* berbasis *Flask*, koneksi internet, serta database *MySQL*. Pengguna dapat memasukkan data faktor lingkungan dan nutrisi tanah melalui antarmuka web, kemudian sistem memproses data tersebut menggunakan model *Random Forest* dan menampilkan hasil rekomendasi tanaman secara *real-time*.

Integrasi antara *Flask* dan *MySQL* memungkinkan sistem untuk mencatat setiap proses prediksi yang dilakukan pengguna. Dengan demikian, sistem tidak hanya memberikan rekomendasi secara *real-time*, tetapi juga menyimpan data hasil prediksi yang dapat digunakan untuk analisis lebih lanjut oleh peneliti atau ahli agronomi.

Halaman Beranda



Gambar 7. Halaman Beranda

Halaman beranda menampilkan identitas sistem dan deskripsi singkat mengenai tujuan serta metode yang digunakan. Sistem ini dirancang untuk membantu pengguna, khususnya petani dan ahli agronomi, dalam memperoleh rekomendasi jenis tanaman berdasarkan kondisi lahan dan faktor lingkungan.

Tampilan ini juga mencantumkan informasi tentang algoritma utama yang digunakan, yaitu *Random Forest*, beserta penjelasan singkat mengenai peran teknologi pembelajaran mesin dalam meningkatkan efisiensi dan akurasi rekomendasi tanaman.

Halaman Formulir Rekomendasi

Gambar 8. Halaman Formulir Rekomendasi

Halaman ini merupakan inti dari sistem rekomendasi tanaman. Pengguna dapat memasukkan data seperti suhu, kelembapan, kecepatan angin, pH tanah, kandungan *nitrogen* (N), *fosfor* (P), *kalium* (K), serta jenis tanah. Setelah semua data diisi, sistem akan mengirimkan data tersebut ke model *Random Forest* yang telah dilatih sebelumnya untuk melakukan prediksi.

Proses input data dilengkapi dengan validasi otomatis agar nilai yang dimasukkan sesuai dengan format numerik

yang dibutuhkan oleh model. Data yang telah dimasukkan juga secara otomatis disimpan dalam *database MySQL*, memungkinkan pengguna untuk melihat kembali riwayat prediksi yang telah dilakukan.

Halaman Hasil Rekomendasi



Gambar 9. Halaman Hasil Rekomendasi

Setelah pengguna mengirimkan data melalui formulir, sistem akan menampilkan hasil prediksi berupa jenis tanaman yang direkomendasikan. Tampilan ini menampilkan dua tanaman dengan probabilitas tertinggi beserta skor prediksi masing-masing.

Hasil ini diperoleh dari model *Random Forest* yang memproses data faktor lingkungan dan nutrisi tanah menggunakan parameter yang telah dioptimasi. Setiap hasil prediksi juga disimpan ke dalam *database MySQL* bersama data input pengguna, sehingga dapat digunakan untuk pelacakan hasil atau evaluasi performa sistem pada periode tertentu.

Halaman Riwayat Prediksi

ID	Nama User	Suhu	Kelembapan	Angin	pH	Nitrogen	Fosfor	Kalsium	Jenis Tanah	Tanggal	Model	Rekomendasi1	Skor1	Rekomendasi2	Skor2	Rekomendasi3	Skor3	Waktu Input
0	Admin	30.0	80.0	10.0	6.5	80.0	50.0	40.0	Sandy	2025-10-10	RandomForest	Padi	32.0%	Gandum	19.0%	Tebu	14.0%	2025-10-10 04:05

Gambar 10. Halaman Riwayat

Halaman ini menampilkan daftar riwayat hasil prediksi yang tersimpan di dalam *database MySQL*. Setiap entri berisi informasi berupa nama pengguna, tanggal prediksi, nilai faktor lingkungan dan nutrisi tanah yang dimasukkan, serta hasil tanaman yang direkomendasikan oleh sistem.

Fitur ini berfungsi untuk memantau konsistensi hasil rekomendasi dari waktu ke waktu dan memungkinkan pengguna melakukan analisis terhadap kondisi lahan tertentu yang sering menghasilkan tanaman tertentu. Dengan penyimpanan data terpusat di *MySQL*, sistem ini tidak hanya memberikan rekomendasi instan tetapi juga berfungsi sebagai sistem pendukung keputusan berbasis data historis.

INTEGRASI SISTEM

Secara keseluruhan, integrasi antara *Flask*, *MySQL*, dan *model Random Forest* membentuk sistem rekomendasi tanaman yang interaktif, efisien, dan berkelanjutan. Alur data dimulai dari input pengguna melalui antarmuka web, kemudian diproses oleh model *Machine Learning* untuk menghasilkan rekomendasi tanaman, dan hasil akhirnya disimpan dalam basis data *MySQL* untuk keperluan pemantauan maupun evaluasi.

Pendekatan ini tidak hanya mempercepat proses rekomendasi, tetapi juga memungkinkan sistem digunakan dalam skala luas tanpa kehilangan akurasi dan konsistensi hasil prediksi.

SIMPULAN

Berdasarkan hasil penelitian dan analisis yang telah dilakukan, dapat disimpulkan beberapa hal penting terkait perbandingan algoritma *Random Forest* dan *Support Vector Machine (SVM)* dalam sistem rekomendasi jenis tanaman berbasis faktor lingkungan dan nutrisi tanah sebagai berikut:

1. Penelitian ini berhasil mengimplementasikan sistem rekomendasi jenis tanaman berbasis *Machine Learning* dengan menggunakan dua algoritma utama, yaitu *Random Forest* dan *Support Vector Machine (SVM)*. Dataset yang digunakan mencakup faktor lingkungan (suhu, kelembapan, kecepatan angin), nutrisi tanah (pH, nitrogen, fosfor, dan kalium), serta jenis tanah yang berperan

- penting dalam menentukan kesesuaian tanaman terhadap kondisi tersebut.
2. Hasil evaluasi menunjukkan bahwa algoritma *Random Forest* memiliki performa terbaik dengan akurasi sebesar 89,03%, sedikit lebih tinggi dibandingkan *SVM* sebesar 88,51%. *Random Forest* juga lebih stabil dalam mengklasifikasikan sebagian besar jenis tanaman, meskipun masih terdapat kesalahan pada kelas gandum dan kentang akibat kemiripan pola data. Hal ini mengindikasikan bahwa *Random Forest* lebih adaptif terhadap variasi data yang kompleks dan *non-linear*.
 3. Analisis *feature importance* menunjukkan bahwa *Nitrogen* (N), *Kalium* (K), dan *Fosfor* (P) merupakan fitur yang memiliki tingkat kepentingan tertinggi dalam menentukan hasil rekomendasi jenis tanaman. Ketiga unsur hara makro tersebut berpengaruh dominan terhadap keputusan sistem, diikuti oleh variabel lingkungan seperti kecepatan angin (*Wind Speed*) dan suhu (*Temperature*) yang memiliki pengaruh sedang, serta pH tanah (*Soil pH*), jenis tanah (*Soil Type*), dan kelembapan (*Humidity*) yang memberikan kontribusi relatif kecil. Hasil ini menegaskan bahwa kondisi nutrisi tanah memiliki peran utama dalam menentukan kecocokan tanaman terhadap lahan.
 4. Implementasi sistem berbasis web menggunakan *Flask* dan *MySQL* berhasil mengintegrasikan model *Random Forest* ke dalam aplikasi interaktif yang mudah digunakan. Sistem ini memungkinkan pengguna untuk memperoleh rekomendasi tanaman secara *real-time* berdasarkan input faktor lingkungan dan nutrisi tanah, serta menyimpan riwayat prediksi untuk analisis lebih lanjut sebagai alat bantu pengambilan keputusan bagi petani maupun ahli agronomi.
- Adaga, & Donald Obinna Daraojimba. (2024). AI in precision agriculture: A review of technologies for sustainable farming practices. *World Journal of Advanced Research and Reviews*, 21(1), 2276–2285.
- Ali, Z., Muhammad, A., Lee, N., Waqar, M., & Lee, S. W. (2025). Artificial Intelligence for Sustainable Agriculture: A Comprehensive Review of AI-Driven Technologies in Crop Production. *Sustainability (Switzerland)*, 17(5), 1–24.
- Anonymous. (2025). Krisis Kesuburan Tanah di Indonesia: Fakta atau Alarm Palsu? [alatrisetpertanian.com](https://www.alatrisetpertanian.com/knowledge/krisis-kesuburan-tanah-di-indonesia-fakta-atau-alarm-palsu/?utm_source) Diakses 8 Oktober 2025, pukul 12.00 dari https://www.alatrisetpertanian.com/knowledge/krisis-kesuburan-tanah-di-indonesia-fakta-atau-alarm-palsu/?utm_source
- Badan Pangan Nasional. (2023). *Rencana Aksi Badan Pangan Nasional Tahun 2023*. Jakarta: Badan Pangan Nasional Republik Indonesia.
- Burla, U. T., Rajan, K., Sanjay, P., Chirumamilla, S., Ravi, G. M., & Ashhar, A. (2024). Advanced Crop Recommendation Systems: Leveraging Random Forest and KNN Algorithms. *International Journal For Multidisciplinary Research*, 6(4), 1–12.
- Dachi, J. M. A. S. D., & Sitompul, P. (2023). Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit. *Jurnal Riset Rumpun Matematika Dan Ilmu Pengetahuan Alam*, 2(2), 87–103.
- FAO. (2024). *The State of Food and Agriculture: Rural transformation and agri-food systems*. Rome: Food and Agriculture Organization.
- Hamdani, I. M., Nurhidayat, N., Karman, A., Adhalia H, N. F., & Julyaningsih, A. H. (2024). Edukasi dan Pelatihan Data Science dan Data Preprocessing.

DAFTAR PUSTAKA

Adewusi, A., Onyeka Franca Asuzu, Temidayo Olorunsogo, Ejuma

- Intisari: Jurnal Inovasi Pengabdian Masyarakat*, 2(1), 19–26.
- Herman. (2022). Dirjen FAO: Digitalisasi Masa Depan Pertanian Dunia. *sinartani.co.id*. Diakses 8 Oktober 2025 pukul 12.30 dari https://sinartani.co.id/nasional/dirjen-fao-digitalisasi-masa-depan-pertanian-dunia/?utm_source
- Kencana, D. (2025). Pentingnya Uji Tanah, Solusi Pemupukan Tepat bagi Petani. *Jateng.Idntimes.Com*. Diakses 8 Oktober 2025 pukul 13.30 dari https://jateng.idntimes.com/news/business/pentingnya-uji-tanah-solusi-pemupukan-tepat-bagi-petani-00-2m8jx-mrrdv8?utm_source
- Kurniawan, R., Halim, A., & Melisa, H. (2023). Prediksi Hasil Panen Pertanian Salak di Daerah Tapanuli Selatan Menggunakan Algoritma SVM (Support Vector Machine). *Kajian Ilmiah Informatika Dan Komputer*, 4(2), 903–912.
- Occelli, M., Mantino, A., Ragaglini, G., Dell’Acqua, M., Fadda, C., Pè, M. E., & Nuvolari, A. (2021). Traditional knowledge affects soil management ability of smallholder farmers in marginal areas. *Agronomy for Sustainable Development*, 41(1).
- Rahman, A., Udjulawa, D., & Mulyati, M. (2024). Rekomendasi Pemilihan Jenis Tanaman Menggunakan Algoritma Random Forest dan XGBoost Regressor. *Computer Science (CO-SCIENCE)*, 4(2), 119–126.
- Sekretariat Jendral. (2023). *Analisis Komoditas Pangan Strategis Tahun 2023*. Jakarta: Kementerian Pertanian Republik Indonesia.
- Senapaty, M. K., Ray, A., & Padhy, N. (2024). A Decision Support System for Crop Recommendation Using Machine Learning Classification Algorithms. *Agriculture (Switzerland)*, 14(8).
- Siregar, S. A., & Nasution, Y. R. (2024). Prediction of Rice Farming Yields in Padangsidempuan City through Support Vector Machine (SVM) Algorithms. *JINAV: Journal of Information and Visualization*, 5(1), 146–156.
- Yasin, F., Firmansyah, M. R., Aldo, D., & Amrustian, M. A. (2025). Multivariate Forecasting of Paddy Production: A Comparative Study of Machine Learning Models. *Jurnal Teknik Informatika (Jutif)*, 6(3), 1431–1442.
- Zalvadila, A., Purnawansyah, Syafie, L., & Darwis, H. (2023). Klasifikasi Penyakit Tanaman Bawang Merah Menggunakan Metode SVM dan CNN. *Jurnal Informatika: Jurnal Pengembangan IT*, 8(3), 255–260.