

Perbandingan Algoritma Klasifikasi untuk Rekomendasi Tanaman Berdasarkan Data Lingkungan

Willy Muhammad Fauzi^{1*}, Adi Dadan Ramdana², Fajar Firmansyah³, Fajar Yudha Nugraha⁴, Akpil Mauhib⁵

^{1,2,3,4,5} Jurusan Teknik Informatika, Fakultas Teknik, Universitas Mayasari Bakti, Tasikmalaya
^{1,2,3,4,5} Jalan Tamansari Blok Rahayu I, RT.004 / RW.004, Kel. Sukahurip, Kec. Tamansari, Kota Tasikmalaya
email: ¹wilymuhf@mail.com, ²adidadanr@gmail.com

Abstract — *Selecting the right crops is crucial for improving agricultural productivity. With machine learning technology, crop recommendation based on environmental data can become more efficient, especially in varying climatic conditions. This study aims to compare three classification algorithms, namely Random Forest, XGBoost, and SVM, in providing crop recommendations based on environmental data including temperature, humidity, soil pH, and rainfall. This research uses a dataset that includes environmental features from BPS Kota Tasikmalaya, which were then tested with three machine learning classification algorithms Random Forest, XGBoost, and SVM. Each model is evaluated based on accuracy, precision, recall, and F1-score.*

Random Forest achieved the best performance with 99.32% accuracy, followed by XGBoost with 98.64%, and SVM with 96.82%. These models recommend crops such as orange and melon, while SVM predominantly recommends mothbeans.

Random Forest provides the most optimal results in the crop recommendation system based on environmental data, although SVM is faster in training the model. This study demonstrates the importance of applying machine learning algorithms to support agricultural decision-making based on local environmental conditions.

Keywords - *Crop Recommendation, Machine Learning, Random Forest, XGBoost, SVM.*

Abstrak Pemilihan tanaman yang tepat sangat penting dalam meningkatkan produktivitas pertanian. Dengan adanya teknologi machine learning, proses rekomendasi tanaman berdasarkan data lingkungan dapat lebih efisien, terutama dalam menghadapi kondisi iklim yang bervariasi. Penelitian ini bertujuan untuk membandingkan tiga algoritma klasifikasi, yaitu Random Forest, XGBoost, dan SVM, dalam memberikan rekomendasi tanaman yang sesuai berdasarkan data lingkungan yang mencakup suhu, kelembaban, pH tanah, dan curah hujan. Penelitian ini menggunakan dataset yang mencakup fitur lingkungan dari BPS Kota Tasikmalaya, yang kemudian diuji dengan tiga algoritma klasifikasi machine learning Random Forest, XGBoost, dan SVM. Setiap model dievaluasi berdasarkan akurasi, precision, recall, dan F1-score. Random Forest menunjukkan hasil terbaik dengan akurasi 99.32%, diikuti oleh XGBoost dengan akurasi 98.64%, dan SVM dengan akurasi 96.82%. Model-model ini memberikan rekomendasi tanaman seperti jeruk dan melon, sementara SVM lebih sering merekomendasikan mothbeans. Random Forest memberikan hasil yang paling optimal dalam sistem rekomendasi tanaman berbasis data lingkungan, meskipun SVM lebih cepat dalam hal pelatihan model. Penelitian ini menunjukkan pentingnya penerapan algoritma machine learning untuk mendukung keputusan pertanian berbasis kondisi lingkungan lokal.

Kata Kunci – *Rekomendasi Tanaman, Machine Learning, Random Forest, XGBoost, SVM.*

*) **penulis korespondensi:** Willy Muhammad Fauzi
Email: wilymuhf@gmail.com

I. PENDAHULUAN

Rekomendasi tanaman menggunakan algoritma machine learning menjadi topik yang semakin penting dalam pertanian modern. Di Indonesia, dengan keanekaragaman iklim dan jenis tanah, penerapan teknologi ini dapat membantu petani dalam memilih tanaman yang paling sesuai untuk meningkatkan hasil panen dan efisiensi penggunaan sumber daya. Algoritma machine learning menawarkan solusi berbasis data yang dapat mengoptimalkan keputusan pertanian dengan mempertimbangkan berbagai faktor lingkungan seperti kondisi tanah dan iklim [1], [2], [3]. Salah satu masalah utama dalam pertanian di Indonesia adalah ketidakpastian dalam memilih jenis tanaman yang tepat untuk ditanam, yang sering kali disebabkan oleh kurangnya informasi yang akurat mengenai kondisi tanah dan iklim. Hal ini dapat mengakibatkan hasil panen yang tidak optimal dan pemborosan sumber daya. Selain itu, perubahan iklim dan pertumbuhan populasi menambah tekanan pada sektor pertanian untuk meningkatkan produktivitas secara berkelanjutan [4], [5].

Penelitian sebelumnya telah menunjukkan bahwa algoritma machine learning seperti Random Forest, Naive Bayes, dan Support Vector Machines dapat digunakan untuk merekomendasikan tanaman berdasarkan data lingkungan [1], [6], [7]. Namun, masih terdapat gap dalam penelitian terkait penerapan spesifik di Indonesia, terutama dalam hal adaptasi algoritma terhadap kondisi lokal dan integrasi data real-time dari sensor IoT [5], [8]. Beberapa studi juga menyoroti pentingnya pemilihan algoritma yang tepat dan teknik pra-pemrosesan data untuk mencapai kinerja optimal [4], [9].

Penelitian ini bertujuan untuk membandingkan berbagai algoritma machine learning dalam konteks rekomendasi tanaman di Indonesia, dengan fokus pada adaptasi terhadap kondisi lokal. Dengan demikian, penelitian ini diharapkan dapat memberikan panduan bagi petani dan pembuat kebijakan dalam memilih teknologi yang tepat untuk meningkatkan produktivitas pertanian. Selain itu, penelitian ini juga berkontribusi pada pengembangan sistem pertanian presisi yang lebih efisien dan berkelanjutan.

Penelitian ini memberikan kontribusi yang signifikan terhadap pengembangan sistem rekomendasi tanaman berbasis data lingkungan menggunakan teknik machine learning, khususnya dalam membandingkan tiga algoritma klasifikasi populer, yaitu Random Forest, XGBoost, dan SVM. Salah satu kontribusi utama dari penelitian ini adalah perbandingan langsung antara ketiga algoritma tersebut dalam konteks aplikasi pertanian, yang jarang ditemukan pada penelitian sebelumnya yang umumnya lebih fokus pada satu atau dua

algoritma saja. Dengan menganalisis kinerja model-model tersebut menggunakan berbagai metrik evaluasi seperti akurasi, precision, recall, dan F1-score, penelitian ini memberikan gambaran yang lebih lengkap mengenai efektivitas dan kehandalan masing-masing algoritma dalam memprediksi jenis tanaman yang sesuai dengan kondisi lingkungan tertentu. Selain itu, penelitian ini juga berkontribusi dalam meningkatkan pemahaman mengenai pentingnya fitur dalam menentukan rekomendasi tanaman. Dengan memvisualisasikan feature importance dari masing-masing model, penelitian ini tidak hanya memberikan akurasi dalam prediksi, tetapi juga membantu mengidentifikasi faktor-faktor lingkungan yang paling berpengaruh dalam memilih jenis tanaman yang optimal. Hal ini tentunya akan memberikan nilai lebih bagi petani atau praktisi pertanian yang ingin membuat keputusan berbasis data mengenai jenis tanaman yang akan dibudidayakan.

II. PENELITIAN YANG TERKAIT

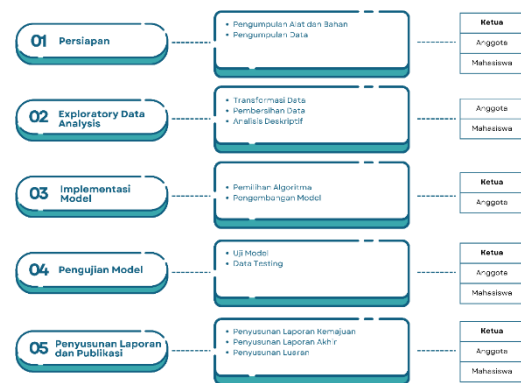
Sistem rekomendasi tanaman berbasis machine learning telah berkembang pesat dalam beberapa tahun terakhir. Beberapa penelitian telah menunjukkan keberhasilan penggunaan algoritma machine learning seperti Random Forest, XGBoost, dan SVM untuk memberikan rekomendasi tanaman yang optimal berdasarkan data lingkungan dan tanah. Garg et al. (2021) mengembangkan sistem rekomendasi tanaman dengan menggunakan pendekatan ensemble yang menggabungkan Random Forest dan XGBoost, memberikan hasil yang akurat dalam memprediksi tanaman yang sesuai dengan kondisi iklim dan tanah lokal di India [10]. Penelitian lain oleh Soni et al. (2019) menerapkan SVM dalam sistem rekomendasi tanaman, dengan hasil yang menunjukkan bahwa SVM efektif dalam memberikan rekomendasi tanaman berdasarkan data iklim dan pertanian, terutama ketika data memiliki batasan yang jelas antar kelas tanaman [11]. Di sisi lain, penelitian oleh Mohapatra dan Kale (2021) menggunakan Random Forest dan SVM dalam rekomendasi tanaman, menunjukkan bahwa Random Forest unggul dalam memprediksi tanaman yang sesuai dengan data lingkungan yang kompleks [12].

Pendekatan yang menggabungkan informasi tanah dengan machine learning juga banyak diuji. Sebagai contoh, Soni et al. (2025) mengintegrasikan data tanah dalam sistem rekomendasi tanaman menggunakan Random Forest dan XGBoost, yang menunjukkan peningkatan akurasi dalam memprediksi tanaman yang optimal [13]. Behera (2024) menunjukkan bagaimana Random Forest dapat memberikan prediksi yang akurat untuk hasil tanaman dengan mempertimbangkan variabel lingkungan seperti curah hujan dan suhu tanah [14]. Di sisi lain, Kumar dan Gupta (2022) menguji XGBoost untuk memprediksi hasil tanaman yang optimal, dengan hasil yang menunjukkan bahwa XGBoost memberikan akurasi yang sangat tinggi dalam memprediksi hasil pertanian berdasarkan data iklim dan kelembaban [15].

Selain itu, beberapa penelitian juga mengadopsi penggunaan data penginderaan jauh (remote sensing) untuk meningkatkan kualitas klasifikasi tanaman. Mahesh et al. (2019) menggunakan XGBoost dalam klasifikasi tanaman berbasis citra satelit, memberikan hasil yang lebih baik dibandingkan dengan model tradisional [16].

Secara keseluruhan, meskipun banyak penelitian sebelumnya yang telah membahas penggunaan machine learning dalam rekomendasi tanaman, penelitian ini memberikan kontribusi dengan membandingkan tiga algoritma klasifikasi utama (Random Forest, XGBoost, dan SVM) dengan evaluasi yang lebih mendalam pada rekomendasi tanaman serta data uji menggunakan data lokal dari BPS Kota Tasikmalaya. Penelitian tidak hanya memperluas pemahaman tentang bagaimana masing-masing algoritma bekerja dalam konteks data lokal, tetapi juga menunjukkan pengaruh nyata dari data lingkungan spesifik regional terhadap hasil rekomendasi tanaman.

III. METODE PENELITIAN



Gbr 1. Metode Penelitian

A. Persiapan

1) Pengumpulan Alat dan Bahan

Perangkat keras dan perangkat lunak yang diperlukan untuk eksperimen, termasuk komputer dengan spesifikasi yang memadai dan perangkat lunak pemrograman seperti Python.

2) Pengumpulan Data

Pengumpulan data dilakukan dengan cara mencari data primer dan data sekunder.

B. Exploratory Data Analysis (EDA)

Tahapan kedua meliputi transformasi dan pembersihan data, visualisasi data, analisis deskriptif serta pengembangan hipotesis. Pada tahapan ini data diolah untuk bisa dilanjutkan ke tahap implementasi model.

C. Implementasi Model

1) Pemilihan Algoritma

Pada tahapan ini penulis akan memilih beberapa algoritma machine learning untuk klasifikasi tanah, seperti Decision Tree, Random Forest, k-Nearest Neighbors (k-NN), dan Support Vector Machine (SVM)

2) Pengujian Model

Pengujian model yang telah dikembangkan dengan data uji untuk mengevaluasi akurasi dan kinerja model. Penulis akan membandingkan hasil akurasi dari setiap algoritma untuk menentukan algoritma yang paling efektif.

D. Analisis Perbandingan Algoritma

Algoritma yang telah ditraining dan dievaluasi akan dianalisis untuk mengetahui algoritma mana yang paling baik dari berbagai macam faktor

IV.HASIL DAN PEMBAHASAN

A. Persiapan

Pengumpulan data dilakukan dengan cara mencari data primer dan data sekunder. Data Primer adalah data kondisi tanah yang ada di tasikmalaya. Peneliti menggunakan data yang berisi Humidity, Temperature dan Curah Hujan dari BPS Kota Tasikmalaya Tahun 2023 (<https://tasikmalayakota.bps.go.id>).

Sedangkan Data Sekunder akan diambil dari Kaggle dengan judul “Crop Recommendation Dataset” yang diupload oleh Athaya Ingle dipublis pada tahun 2021 yang berisi N,P,K, Suhu, Kelembapan dan Curah Hujan.

Tabel 1 Fitur Dataset Crop Recommendation

No	Fitur	Deskripsi
1	N	Kandungan Nitrogen dalam Tanah
2	P	Kandungan Fosfor dalam tanah
3	K	Kandungan Kalium dalam tanah
4	Temperature	Suhu rata-rata dalam lingkungan tempat tanaman tumbuh
5	Humidity	Kelembapan udara di sekitar tanaman
6	pH	Tingkat Keasaman Tanah
7	Rainfall	Jumlah Curah Hujan

Data “Crop Recommendation Dataset” akan dijadikan data training dan data testing untuk melakukan perbandingan model menggunakan algoritma yang telah ditentukan yaitu Random Forest, XGBoost dan Super Vector Machine (SVM). Selanjutnya data lokal yang berasal dari BPS Kota Tasikmalaya akan dijadikan data Uji untuk melihat hasil rekomendasi yang memungkinkan hasil lebih relevan untuk wilayah tersebut.

B. EDA (Exploratory Data Analysis)

Pada bagian ini data yang kita punya akan dianalisis untuk mendapatkan pemahaman yang lebih baik mengenai data yang kita punya.

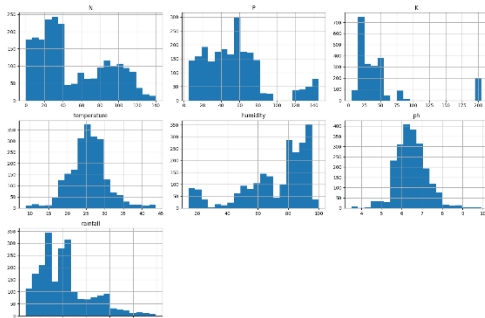
1) Deskripsi Data

Dataset yang digunakan dalam penelitian ini berjumlah 2200 entri dengan 8 fitur, yaitu: N, P, K, temperature, humidity, ph, rainfall, dan label (target variabel yang berisi jenis tanaman). Seluruh data bersifat numerik kecuali kolom label yang berupa kategori (jenis tanaman).

2) Statistik Deskriptif

Berdasarkan statistik deskriptif, fitur-fitur seperti N, P, dan K menunjukkan variasi yang cukup signifikan, dengan nilai maksimum pada K yang jauh lebih tinggi daripada nilai kuartil ketiga. Fitur temperature, humidity, dan rainfall memiliki sebaran nilai yang lebih terkonsentrasi, dengan rata-rata suhu sekitar 25°C dan kelembapan sekitar 71%.

3) Visualisasi Data

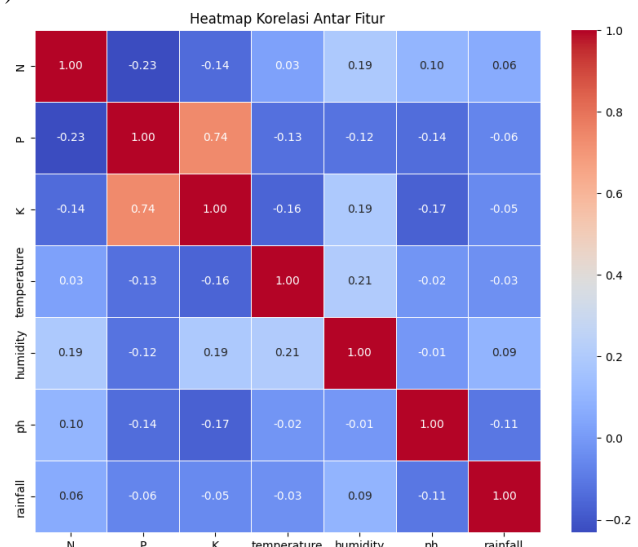


Gbr 2. Visualisasi Data

Berikut adalah penjelasan untuk setiap fitur yang divisualisasikan:

- Histogram untuk N menunjukkan distribusi yang agak miring dengan beberapa puncak. Sebagian besar data terkonsentrasi pada nilai antara 0 hingga 50, dengan beberapa nilai yang lebih tinggi di kisaran 90 hingga 140. Ini menunjukkan adanya skewness (kemiringan data), yang dapat mempengaruhi model jika tidak ditangani dengan baik.
- Distribusi P memiliki puncak tajam di sekitar nilai 60-70, dengan sebagian besar data terkonsentrasi di kisaran ini. Beberapa nilai juga lebih tinggi, tetapi jumlahnya relatif sedikit, yang menunjukkan adanya outliers (nilai pencilan) yang perlu diperhatikan lebih lanjut
- Pada fitur K, kita melihat adanya skewness yang sangat kuat, di mana sebagian besar data terkonsentrasi pada nilai rendah (0-25), dan ada puncak yang cukup signifikan pada angka sekitar 25. Nilai lebih tinggi dari 50 jarang ditemui, menunjukkan bahwa ada kemungkinan outliers pada fitur ini, yang dapat mempengaruhi analisis.
- Distribusi temperature terlihat relatif normal, dengan puncak tertinggi pada kisaran 20-25°C, yang merupakan rentang suhu yang umum untuk pertanian tropis. Data tersebar lebih merata dengan beberapa nilai ekstrim yang menunjukkan adanya variasi suhu.
- Humidity menunjukkan distribusi yang skewed ke kanan, dengan sebagian besar data terkonsentrasi di kisaran 70 hingga 100%. Histogram ini menunjukkan bahwa kelembapan cenderung lebih tinggi pada sebagian besar data.
- Fitur ph menunjukkan distribusi yang mirip dengan normal distribution dengan puncak di sekitar nilai pH 6, yang merupakan pH ideal untuk banyak jenis tanaman. Data tersebar cukup merata di sekitar nilai tersebut, menunjukkan bahwa pH tanah relatif stabil.
- Distribusi rainfall terlihat miring, dengan sebagian besar data terkonsentrasi pada nilai curah hujan rendah hingga menengah (50-150 mm). Namun, ada beberapa data yang menunjukkan curah hujan tinggi, dengan outliers yang cukup signifikan di sisi kanan histogram.

4) Analisis Korelasi Antar Fitur



Gbr 3. EDA – Heatmap Korelasi

Berdasarkan analisis heatmap korelasi pada gambar 3 menunjukkan hubungan antar fitur dalam dataset. Dari analisis, terlihat bahwa P (Phosphorus) dan K (Potassium) memiliki korelasi positif yang kuat (0.74), menunjukkan bahwa keduanya cenderung berfluktuasi bersama. Temperature dan humidity juga memiliki korelasi positif yang moderat (0.21), menandakan hubungan antara suhu dan kelembaban. Namun, fitur seperti ph (pH tanah) memiliki korelasi yang sangat lemah dengan fitur lainnya, menunjukkan bahwa pH tanah tidak dipengaruhi secara signifikan oleh faktor-faktor lain dalam dataset ini. Secara keseluruhan, heatmap ini membantu memahami hubungan antar fitur dan dapat memberikan wawasan dalam memilih fitur yang relevan untuk pemodelan.

5) Distribusi Target Variabel

Distribusi label target menunjukkan bahwa setiap jenis tanaman memiliki jumlah yang seimbang, dengan 100 entri untuk masing-masing label tanaman, seperti rice, maize, dan chickpea. Oleh karena itu, dataset ini tidak mengalami masalah ketidakseimbangan data pada target variabel.

6) Pemeriksaan Missing Values

Pemeriksaan terhadap missing values menunjukkan bahwa tidak ada data yang hilang pada fitur mana pun. Semua fitur memiliki jumlah entri yang lengkap, sehingga tidak diperlukan imputasi data lebih lanjut.

C. Inisialisasi Data

Data dibagi menjadi dua bagian utama yaitu data pelatihan (training data) dan data pengujian (test data).

```
[ ] # Membagi data menjadi training dan testing
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42)
```

Gbr 4. Pembagian Data Set

Pembagian ini dilakukan dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian, serta menggunakan parameter random_state=42 untuk memastikan hasil yang dapat direproduksi.

D. Implementasi Model

Pada tahap ini akan dilakukan implementasi model menggunakan 3 algoritma klasifikasi yaitu Random Forest, XGBoost dan SVM.

1) Random Forest

a. Training Model

Model Random Forest dibuat dengan menggunakan RandomForestClassifier dari pustaka sklearn.ensemble. Parameter random_state=42 digunakan untuk memastikan hasil yang konsisten pada setiap eksperimen.

```
import time
# Mulai timer
start_time = time.time()

# Model Random Forest
rf = RandomForestClassifier(random_state=42)
rf.fit(X_train, y_train)

# Akhir timer
end_time = time.time()

# Waktu yang diperlukan untuk training (dalam detik)
training_time = end_time - start_time
print(f"Training time for Random Forest: {training_time:.4f} seconds")
```

Gbr 5. Inisialisasi Model RF

Model dilatih menggunakan data pelatihan dengan memanggil fungsi fit() pada objek model. Proses pelatihan ini membangun beberapa pohon keputusan yang masing-masing membuat prediksi berdasarkan subset acak dari data dan fitur. Setiap pohon belajar dari data yang berbeda dan memberikan "suara" untuk hasil akhir. Random Forest kemudian memilih hasil yang paling sering dipilih oleh pohon-pohon ini.

Model Random Forest berhasil dilatih dalam waktu yang sangat cepat, yaitu 0.416 detik, yang menunjukkan efisiensi algoritma dalam menangani dataset besar dengan berbagai fitur. Kecepatan pelatihan ini sangat penting ketika menerapkan model dalam skala besar atau dalam situasi yang membutuhkan waktu pemrosesan yang singkat

b. Evaluasi Model

Setelah model dilatih, selanjutnya melakukan prediksi menggunakan data pengujian (test data) pada model yang telah dilatih. Kinerja model dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score. Selain itu, confusion matrix juga digunakan untuk melihat lebih jelas performa model dalam mengklasifikasikan setiap kelas tanaman.

Tabel 2. Evaluasi Metrik Random Forest

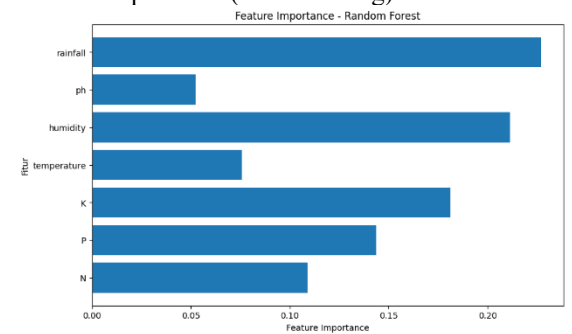
Accuracy	Precision	Recall	F1-Score
0.993182	0.993735	0.993182	0.993175

c. Hasil Analisis

- Akurasi dan Kinerja Model

Berdasarkan hasil yang diperoleh dari pengujian, model Random Forest menunjukkan akurasi sebesar 99.32%, precision 99.37%, recall 99.32%, F1-Score 99.32% yang menunjukkan bahwa model ini sangat efektif dalam mengklasifikasikan jenis tanaman berdasarkan fitur yang diberikan.

- Feature Importance (Fitur Penting)



Gbr 6. Fitur Penting RF

Gambar 6 menunjukkan bahwa rainfall (curah hujan) memiliki pengaruh terbesar terhadap prediksi model Random Forest dengan nilai importance sekitar 0.21, diikuti oleh humidity (kelembaban) dan temperature (suhu) yang masing-masing memiliki pengaruh yang signifikan. Fitur pH tanah dan K (kalium) menunjukkan kontribusi yang lebih kecil, dengan nilai importance di bawah 0.10. Ini mengindikasikan bahwa curah hujan, suhu, dan kelembaban memiliki peran utama dalam rekomendasi tanaman, sementara pH tanah dan kalium kurang berpengaruh dalam model ini.

2) XGBoost

a. Training Model

Model diinisialisasi dan dilatih menggunakan data pelatihan dengan fungsi fit(). Proses pelatihan ini bertujuan untuk membangun banyak pohon keputusan (decision trees) yang bekerja bersama untuk meningkatkan akurasi.

```
import time
# Mulai timer
start_time = time.time()

# Membuat model XGBoost
xgboost_model = xgb.XGBClassifier(random_state=42)

# Melatih model menggunakan data training
xgboost_model.fit(X_train, y_train)

# Akhir timer
end_time = time.time()

# Waktu yang diperlukan untuk training (dalam detik)
training_time = end_time - start_time
print(f"Training time for Xgb: {training_time:.4f} seconds")
```

Training time for Xgb: 1.7547 seconds

Gbr 7. Inisialisasi Model XGBoost

Parameter penting yang digunakan dalam model ini adalah `random_state` untuk memastikan hasil yang dapat direproduksi dan beberapa parameter lainnya yang ditetapkan sesuai dengan data yang ada.

Model XGBoost membutuhkan waktu pelatihan sekitar 1.75 detik, yang lebih lama dibandingkan dengan Random Forest yang hanya membutuhkan 0.416 detik. Meskipun proses pelatihan XGBoost lebih lama, algoritma ini bisa memberikan akurasi yang lebih tinggi berkat teknik gradient boosting yang dapat meningkatkan kualitas prediksi dengan mengurangi kesalahan model secara bertahap.

b. Evaluasi Model

Tabel 2 Evaluasi Metrik XGBoost

Accuracy	Precision	Recall	F1-Score
0.986364	0.986901	0.986364	0.986347

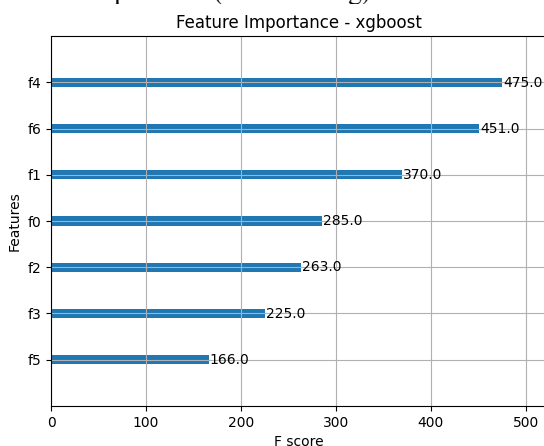
Hasil evaluasi menunjukkan bahwa model XGBoost memiliki performa yang sangat baik dengan akurasi 98.64%, menunjukkan keandalannya dalam melakukan klasifikasi tanaman berdasarkan data yang diberikan.

c. Hasil Analisis

- Akurasi dan Kinerja Model

Semua metrik evaluasi lainnya, seperti precision (98.69%), recall (98.64%), dan F1-score (98.63%) juga menunjukkan bahwa model ini mampu memberikan hasil yang konsisten dan seimbang, dengan sedikit kesalahan dalam memprediksi kelas tanaman yang sesuai.

- Feature Importance (Fitur Penting)



Gbr 8. Fitur Penting XGBoost

Grafik di atas menunjukkan feature importance dari model XGBoost berdasarkan F score. Fitur f4 (memiliki nilai F score tertinggi (475), yang menunjukkan bahwa fitur ini paling berpengaruh dalam keputusan model.

Fitur f6 dan f1 juga menunjukkan kontribusi yang signifikan dengan nilai F score masing-masing 451 dan 370. Sementara itu, fitur f5 memiliki kontribusi terendah dengan F score 166. Hal ini menunjukkan bahwa f4 memiliki peran paling besar dalam prediksi model XGBoost.

3) Support Vector Machine

a. Training Model

Model SVM diinisialisasi menggunakan SVC dari pustaka `sklearn.svm`, dengan menggunakan kernel linear terlebih dahulu untuk memisahkan kelas secara langsung.

```
[ ] import time
# Mulai timer
start_time = time.time()

# Model SVM
svm = SVC(random_state=42)
svm.fit(X_train, y_train)

# Akhir timer
end_time = time.time()

# Waktu yang diperlukan untuk training (dalam detik)
training_time = end_time - start_time
print(f"Training time for SVM: {training_time:.4f} seconds")
```

Training time for SVM: 0.0466 seconds

Gbr 9. Inisialisasi Model SVM

Model dilatih dengan memanggil `fit()` menggunakan data pelatihan. Dalam proses pelatihan, SVM akan menentukan hyperplane terbaik yang memisahkan kelas tanaman.

Proses pelatihan untuk model SVM hanya membutuhkan waktu 0.047 detik, yang jauh lebih cepat dibandingkan dengan Random Forest yang memerlukan waktu pelatihan sekitar 0.416 detik. Meskipun SVM lebih efisien dalam hal waktu komputasi, Random Forest cenderung memberikan hasil yang lebih stabil dan akurat, meskipun membutuhkan lebih banyak waktu pelatihan.

b. Evaluasi Model

Tabel 3 Evaluasi Metrik SVM

Accuracy	Precision	Recall	F1-Score
0.968182	0.971517	0.968182	0.968027

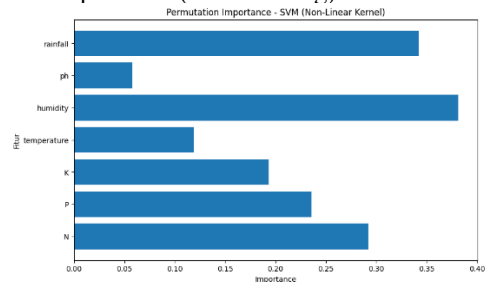
Hasil evaluasi menunjukkan bahwa model SVM memiliki performa yang sangat baik dengan akurasi 96.82%. Akan tetapi akurasi ini lebih rendah dibandingkan dengan algoritma Random Forest dan XGBoost.

c. Hasil Analisis

- Akurasi dan Kinerja Model

Metrik evaluasi lainnya, seperti precision (97.15%), recall (96.82%), dan F1-score (96.80%), juga menunjukkan bahwa model ini mampu memberikan hasil yang konsisten dan seimbang dalam mengklasifikasikan jenis tanaman.

- Feature Importance (Fitur Penting)



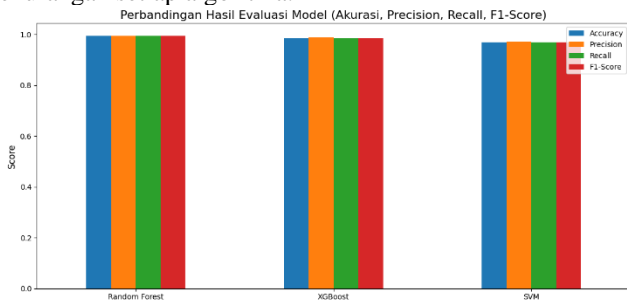
Gbr 9. Fitur Penting SVM

Curah hujan (rainfall) dan kelembaban (humidity) adalah fitur yang paling berpengaruh dalam SVM dengan non-linear kernel untuk rekomendasi tanaman, sementara pH, kalium (K), fosfor (P), dan nitrogen (N) memiliki pengaruh yang lebih kecil dalam model ini.

E. Perbandingan Model

1) Evaluasi Metrik

Pada bagian ini akan dijelaskan secara rinci mengenai perbandingan kinerja dari ketiga model yang sudah digunakan, yaitu Random Forest, XGBoost, dan SVM. Penjelasan mencakup perbandingan akurasi, precision, recall & F1-score serta waktu pelatihan masing-masing model untuk memberikan gambaran yang jelas tentang kelebihan dan kekurangan setiap algoritma.



Gbr 10. Perbandingan Hasil Evaluasi

a. Akurasi

- Random Forest memiliki akurasi 99.32%, yang menunjukkan hasil yang sangat baik dalam memprediksi tanaman yang tepat berdasarkan fitur lingkungan.
- XGBoost memiliki akurasi 98.64%, sedikit lebih rendah dari Random Forest, namun tetap menunjukkan performa yang sangat baik.
- SVM menunjukkan akurasi 96.82%, yang lebih rendah dibandingkan dengan kedua model lainnya, tetapi tetap menunjukkan kinerja yang solid.

b. Precision

- Random Forest memiliki precision 99.37%, yang menunjukkan bahwa prediksi positif dari model lebih jarang salah.
- XGBoost juga memiliki precision yang sangat baik, yaitu 98.69%, menunjukkan bahwa model ini juga jarang membuat kesalahan prediksi.
- SVM memiliki precision 97.15%, yang sedikit lebih rendah, namun masih menunjukkan kinerja yang baik.

c. Recall

- Random Forest dan XGBoost masing-masing memiliki recall 99.32% dan 98.64%, yang menunjukkan bahwa keduanya mampu menangkap hampir semua contoh dari kelas target (jenis tanaman).
- SVM memiliki recall 96.82%, yang sedikit lebih rendah, menunjukkan bahwa model ini melewatkan beberapa contoh dari kelas target.

d. F1-Score

- Random Forest memiliki F1-score 99.32%, menunjukkan keseimbangan yang sangat baik antara precision dan recall.
- XGBoost memiliki F1-score 98.63%, yang sedikit lebih rendah, tetapi tetap sangat baik.

- SVM memiliki F1-score 96.80%, yang menunjukkan keseimbangan yang lebih rendah dibandingkan kedua model lainnya, meskipun masih sangat baik.

2) Waktu Pelatihan Model

- Random Forest membutuhkan waktu pelatihan 0.416 detik.
- XGBoost memerlukan waktu pelatihan 1.75 detik, lebih lama dibandingkan Random Forest.
- SVM sangat cepat dengan waktu pelatihan hanya 0.047 detik, yang menjadikannya pilihan efisien untuk skenario dengan kebutuhan waktu yang terbatas.

3) Kelebihan dan Kekurangan Model

- Random Forest memberikan hasil terbaik dalam semua metrik evaluasi, tetapi membutuhkan waktu pelatihan yang lebih lama dibandingkan SVM. Kelebihan utama adalah kestabilannya dalam mengatasi data yang kompleks
- XGBoost Meskipun memerlukan waktu pelatihan lebih lama, XGBoost menawarkan akurasi dan precision yang sangat baik, mendekati Random Forest, namun tidak selalu memberikan hasil lebih baik secara signifikan
- SVM memiliki waktu pelatihan yang sangat cepat, tetapi akurasi dan recall-nya sedikit lebih rendah. SVM mungkin lebih cocok untuk dataset yang lebih kecil atau ketika waktu pelatihan menjadi faktor kritis.

4) Kesimpulan Perbandingan Model

Berdasarkan hasil perbandingan, Random Forest menunjukkan kinerja terbaik dalam hal akurasi, precision, recall, dan F1-score, namun dengan waktu pelatihan yang lebih lama dibandingkan SVM. XGBoost memberikan hasil yang sangat baik, tetapi tidak sebaik Random Forest dalam hal akurat dan presisi. Sementara SVM memberikan hasil yang lebih cepat dengan kinerja yang sedikit lebih rendah.

Maka dengan ini Random Forest adalah pilihan terbaik dalam hal akurasi dan kestabilan model, tetapi jika waktu pelatihan menjadi faktor yang penting, maka SVM bisa menjadi alternatif yang lebih efisien.

F. Uji Data Lokal

Setelah mengimplementasikan dan melatih model menggunakan SVM, Random Forest, dan XGBoost. Langkah selanjutnya adalah menganalisis hasil prediksi yang diberikan oleh ketiga model tersebut. Hasil prediksi ini penting untuk mengevaluasi kinerja masing-masing algoritma dalam memberikan rekomendasi tanaman berdasarkan data uji lokal dari BPS Kota Tasikmalaya.

Tabel 4 Data Lokal BPS Kota Tasikmalaya

Bulan	Temperature	Humidity	Rainfall
Jan	25.8	85	258
Feb	25.3	84	229
Mar	26	80	293
Apr	25.1	79	223
Mei	25.4	80	131.5
Jun	25.2	80	104.8
Jul	24.5	82	55.3
Agu	24	77	0.5

Sept	24.9	75	1
Okt	27.1	68	8
Nov	25.2	78	134
Des	26.2	80	194

Tabel dibawah ini menyajikan hasil prediksi untuk masing-masing model pada 12 entri data uji. Dalam analisis ini, kita akan membandingkan perbedaan prediksi yang diberikan oleh setiap model dan melihat bagaimana mereka merekomendasikan jenis tanaman berdasarkan fitur lingkungan yang ada.

Tabel 5 Hasil Rekomendasi Tanaman

Bulan	SVM	Random Forest	XGBoost
Jan	mothbeans	orange	orange
Feb	mothbeans	orange	orange
Mar	mothbeans	orange	orange
Apr	mothbeans	orange	orange
Mei	mothbeans	orange	orange
Jun	mothbeans	orange	orange
Jul	mothbeans	orange	orange
Agu	mothbeans	muskmelon	muskmelon
Sept	mothbeans	muskmelon	muskmelon
Okt	mothbeans	mothbeans	muskmelon
Nov	mothbeans	orange	orange
Des	mothbeans	orange	orange

KESIMPULAN

Penelitian ini berhasil menciptakan model klasifikasi berbasis machine learning untuk memberikan rekomendasi tanaman. Model yang digunakan mencakup Random Forest, XGBoost, dan SVM, yang dilatih menggunakan data lingkungan yang mencakup fitur-fitur seperti suhu, kelembaban, curah hujan, dan pH tanah. Setiap model menunjukkan hasil yang baik dalam memprediksi jenis tanaman yang sesuai berdasarkan data yang diberikan, dengan Random Forest memberikan hasil terbaik dalam hal akurasi, precision, recall, dan F1-score

Perbandingan kinerja tiga algoritma klasifikasi menunjukkan bahwa Random Forest memberikan hasil yang paling optimal dengan akurasi 99.32%, diikuti oleh XGBoost dengan akurasi 98.64%, dan SVM dengan akurasi 96.82%. Random Forest unggul dalam hal precision, recall, dan F1-score, yang menandakan keandalannya dalam memberikan rekomendasi tanaman yang tepat dan konsisten. Sementara itu, SVM memberikan hasil yang lebih rendah pada metrik evaluasi, meskipun lebih cepat dalam waktu pelatihan dibandingkan kedua model lainnya.

Berdasarkan data uji yang diambil dari BPS Kota Tasikmalaya, hasil rekomendasi tanaman yang diberikan oleh Random Forest dan XGBoost cenderung merekomendasikan jeruk (orange) dan melon (muskmelon), sementara SVM secara konsisten merekomendasikan kacang polong (mothbeans). Hal ini menunjukkan bahwa model dapat

memberikan rekomendasi tanaman yang sesuai berdasarkan kondisi lingkungan lokal, dengan SVM memberikan rekomendasi yang lebih terfokus pada satu jenis tanaman.

UCAPAN TERIMA KASIH

Terima kasih kepada Universitas Mayasari Bakti yang telah mendanai penelitian ini melalui Lembaga Penelitian dan Pengabdian kepada Masyarakat (No. Kontrak : 012/Kontrak-Penelitian/LPPM/UMB/II/2025)

DAFTAR PUSTAKA

- [1] A. Changela, Y. Kumar, and A. Koul, "Machine Learning-based Approaches for Crop Recommendations and Prediction," in *2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI)*, IEEE, Nov. 2023, pp. 370–376. doi: 10.1109/ICCSAI59793.2023.10421406.
- [2] S. Gambhir, M. Sharma, K. Agarwal, K. Kumar, L. Kumar, and M. Chaudhary, "Crop Recommendation System Using Machine Learning," *2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, pp. 523–530, 2023, doi: 10.1109/ICAC3N60023.2023.10541784.
- [3] Dolli, P. Rawat, M. Bajaj, S. Vats, and V. Sharma, "An Analysis of Crop Recommendation Systems Employing Diverse Machine Learning Methodologies," *2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT)*, pp. 619–624, 2023, doi: 10.1109/DICCT56244.2023.10110085.
- [4] S. Bhatnagar, N. Lakshmi, A. A, and V. S. M. Y, "Development of a Crop Recommendation System Through the Use of Various Machine Learning Algorithms," *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, pp. 1–6, 2024, doi: 10.1109/ic-ETITE58242.2024.10493758.
- [5] E. Elbasi *et al.*, "Crop Prediction Model Using Machine Learning Algorithms," *Applied Sciences*, p., 2023, doi: 10.3390/app13169288.
- [6] U. P. Jakarbet, S. P, S. M, and A. P, "An Empirical Analysis of Machine Learning Algorithms for Agricultural Crop Prediction," *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 1782–1790, 2024, doi: 10.1109/I-SMAC61858.2024.10714677.
- [7] S. R. Sani, S. V. S. Ummadi, S. Thota, N. Muthineni, V. S. S. Swargam, and T. S. Ravella, "Crop Recommendation System using Random Forest Algorithm in Machine Learning," *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, pp. 501–505, 2023, doi: 10.1109/ICAAIC56838.2023.10141384.
- [8] T. Van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Comput. Electron. Agric.*, vol. 177, p. 105709, 2020, doi: 10.1016/j.compag.2020.105709.
- [9] H. Lahza *et al.*, "Optimization of Crop Recommendations Using Novel Machine Learning Techniques," *Sustainability*, p., 2023, doi: 10.3390/su15118836.
- [10] S. J. Jang, H. H. Park, and Y. I. Kuk, "Application of Various Extracts Enhances the Growth and Yield of Cucumber

(*Cucumis sativus* L.) without Compromising the Biochemical Content,” *Agronomy*, vol. 11, no. 3, p. 505, Mar. 2021, doi: 10.3390/agronomy11030505.

[11] S. S. , T. R. K. , R. A. S. Soni, “A machine learning approach to crop recommendations,” *Int J Environ Res*, vol. 19, pp. 148–158, 2019.

[12] B. N. , K. V. Mohapatra, “Crop recommendation system using machine learning algorithms,” *International Journal of Computer Applications*, vol. 9, pp. 12–23, 2021.

[13] “Incorporating soil information with machine learning for crop recommendation,” *Soni, S. R., Tiwari, R. K., Rawat, A. S.*, vol. 25, 2025.

[14] S. Behera, “Smart crop prediction using random forest and machine learning models,” *SSRN Electronic Journal*, vol. 13, pp. 244–257, 2024.

[15] A. , G. R. Kumar, “Predicting optimal crop yields using machine learning algorithms,” *Agricultural Environmental Sciences*, vol. 6, p. 172, 2022.

[16] S. , J. S. S. , P. S. R. Mahesh, “Integration of remote sensing data and machine learning for crop classification,” *Journal Remote Sensing*, vol. 11, 2019.