

## Rekomendasi Pemilihan Jenis Tanaman Menggunakan Algoritma Random Forest dan XGBoost Regressor

Abdul Rahman<sup>1\*</sup>, Daniel Udjulawa<sup>2</sup>, Mulyati<sup>3</sup>

<sup>1,2,3</sup>Universitas Multi Data Palembang  
Jln. Rajawali No. 14 Palembang – Sumatera Selatan, Indonesia

e-mail: [1arahman@mdp.ac.id](mailto:1arahman@mdp.ac.id), [2daniel@mdp.ac.id](mailto:2daniel@mdp.ac.id), [3muliati@mdp.ac.id](mailto:3muliati@mdp.ac.id)

(\* ) Corresponding Author

Artikel Info : Diterima : 03-11-2023 | Direvisi : 05-07-2024 | Disetujui : 09-07-2024

**Abstrak** - Rekomendasi tanaman yang sesuai dengan kondisi lingkungan dan nutrisi tanah tempat tanaman ditanam dapat memberikan hasil panen yang optimal. Penerapan machine learning pada bidang pertanian telah banyak dilakukan terutama untuk meningkatkan hasil panen. Penelitian ini bertujuan untuk membandingkan dua algoritma *machine learning*, yaitu *Random forest* dan *XGBoost Regressor* yang di implementasikan dalam merekomendasikan tanaman yang sesuai dengan kondisi lingkungan dan nutrisi tanah. Implementasi dari kedua algoritma ini dibandingkan tingkat akurasi menggunakan tiga alat ukur akurasi, yaitu *Mean Absolute Error(MAE)*, *Mean Square Error(MSE)*, dan  $R^2$ . Hasil yang didapat menunjukkan kedua algoritma mempunyai tingkat akurasi yang tidak jauh berbeda. Algoritma *random forest* mempunyai tingkat akurasi yang lebih baik menggunakan *MAE* dan *MSE*, yaitu sebesar 36.73681574 dan 1.848396760, sedangkan algoritma *XGBoost Regressor* mempunyai tingkat akurasi yang baik dengan menggunakan alat ukur akurasi  $R^2$  atau R-Square dengan tingkat akurasi sebesar 0.98542963509705.

Kata Kunci : Machine Learning, Random Forest, Rekomendasi Tanaman, XGBoost

**Abstracts** - Recommendations for plants that suit a particular planting location's environmental conditions and soil nutrients can lead to optimal harvest outcomes. Machine learning applications in agriculture have been widely explored, particularly in enhancing crop yields. This research compares two machine learning algorithms, Random Forest and XGBoost Regressor, which recommend crops suitable for environmental conditions and soil nutrition. The implementation of both algorithms was compared in terms of accuracy using three accuracy metrics: Mean Absolute Error (MAE), Mean Square Error (MSE), and  $R^2$ . The results indicated that both algorithms exhibited comparable accuracy levels. The Random Forest algorithm demonstrated superior accuracy in terms of MAE and MSE, with values of 36.73681574 and 1.848396760, respectively. Meanwhile, the XGBoost Regressor algorithm displayed good accuracy, mainly when measured using the  $R^2$  accuracy metric, achieving a high accuracy level of 0.98542963509705..

Keywords : Machine Learning, Random Forest, Crop Recommendation, XGBoost

### PENDAHULUAN

Perkembangan tanaman dipengaruhi oleh empat faktor utama, yaitu: cahaya, air, suhu, dan nutrisi. Keempat faktor ini mempengaruhi aktivitas hormon pertumbuhan tanaman, memacu atau memperlambat proses pertumbuhannya. Hal ini juga disampaikan oleh Layanan Perluasan Universitas Oregon State, yang menyatakan bahwa faktor-faktor tersebut (cahaya, air, suhu, dan ketersediaan nutrisi) merupakan elemen-elemen utama yang paling berpengaruh terhadap perkembangan tanaman (Craker, 2021)(Upadhyay dkk., 2022). Dinamika perubahan iklim dan kondisi tanah sebagai media tanam membuat para petani memerlukan pendekatan yang cerdas dan dapat diandalkan untuk memilih jenis tanaman yang optimal guna mencapai hasil panen yang maksimal. Teknologi dalam bentuk algoritma *machine learning* pada saat ini telah terbukti banyak membantu dalam memberikan rekomendasi yang akurat dan dapat diandalkan (Meshram dkk., 2021). *Machine learning* dengan komputer *vision* dikaji untuk klasifikasi berbagai set gambar tanaman guna memantau kualitas tanaman dan prediksi hasil panen yang meliputi parameter tanah seperti karbon organik dan kandungan kelembaban, deteksi penyakit dan gulma



pada tanaman, serta deteksi spesies (Sharma dkk., 2021). Algoritma machine learning digunakan untuk membangun indikator penilaian iklim terpadu (ICAI), indikator ini untuk wilayah tertentu dapat memberikan dampak komprehensif dari berbagai faktor meteorologi untuk kesesuaian iklim suatu tanaman (X. Xu dkk., 2019). Penelitian-penelitian tersebut memperlihatkan peran penting perkembangan teknologi *machine learning* yang mampu memberikan kemudahan dan kecerdasan di bidang pertanian.

*Machine Learning* dapat diterapkan dalam sektor pertanian untuk memperkirakan parameter tanah seperti kadar karbon organik dan kelembaban, serta memprediksi hasil panen, mengidentifikasi penyakit dan gulma pada tanaman, dan mendeteksi spesies (Sharma dkk., 2021). Faktor-faktor cahaya, air, suhu, dan nutrisi memberikan pengaruh cukup besar pada pertumbuhan tanaman, oleh sebab itu perlu dikaji untuk menentukan jenis tanaman yang sesuai dengan kondisi faktor-faktor tersebut. Para peneliti juga telah banyak melakukan penelitian terkait dengan rekomendasi tanaman pada suatu wilayah berdasarkan faktor-faktor yang mempengaruhi pertumbuhan tanaman tersebut. Algoritma machine learning *Support Vector Machine* (SVM), *Artificial Neural Network* (ANN), *Random Forest* (RF), *Multivariate Linear Regression* (MLR), dan *K-Nearest Neighbour* (KNN) digunakan untuk pemilihan daftar tanaman yang paling menguntungkan atau memprediksi hasil tanaman untuk tanaman yang dipilih oleh pengguna berdasarkan lokasi penanaman, dimana algoritma RF menunjukkan hasil terbaik dengan akurasi 95% (S M Pande dkk., 2021). Algoritma *machine learning* KNN, *Decision Tree*, RF, *Naive Bayes*, dan *Gradient Boosting* digunakan untuk merekomendasikan tanaman berdasarkan kondisi tanah tertentu dimana tanaman akan ditanam, hasilnya algoritma *Gradient Boosting* mempunyai tingkat akurasi terbaik sebesar 98,18% (Shilpa Mangesh Pande dkk., 2021). Sistem rekomendasi tanaman pangan berbasis logika fuzzy untuk menangani delapan tanaman pangan utama yang akan ditanam, kinerja model ini divalidasi oleh kumpulan data yang beragam dan mencapai akurasi sekitar 92% (Banerjee dkk., 2021).

Pada penelitian ini dilakukan pengujian tingkat akurasi rekomendasi tanaman dari dua algoritma *machine learning*, yaitu algoritma RF dan algoritma *XGBoost Regressor*. Pemilihan algoritma RF pada penelitian ini karena algoritma RF mempunyai tingkat akurasi terbaik dibandingkan dengan algoritma SVM, ANN, MLR, dan KNN dimana faktor-faktor yang digunakan untuk merekomendasikan tanaman berdasarkan suhu, lokasi, iklim, tipe tanah, dan curah hujan (S M Pande dkk., 2021), sedangkan algoritma *XGBoost Regressor* mempunyai keunggulan dalam waktu komputasi yang lebih cepat dan kemampuan menangani data secara akurat dengan banyak fitur (Supriya & Akki, 2021). Pada penelitian ini faktor-faktor yang digunakan untuk merekomendasikan jenis tanaman menggunakan algoritma RF dan *XGBoost Regressor* terdiri dari kandungan nutrisi tanah Nitrogen (N), Fosfor (P), dan Kalium (K), suhu, kelembaban, dan pH tanah, serta curah hujan.

Penelitian terhadap prediksi hasil panen pertanian menggunakan algoritma RF dan *XGBoost Regressor* telah banyak dilakukan. Algoritma RF dengan pendekatan berbasis peringkat digunakan untuk untuk prediksi hasil jagung, pendekatan ini memberikan berkontribusi pada pengambilan keputusan, sehingga menghasilkan pengelolaan lahan jagung yang akurat (Marques Ramos dkk., 2020). Pendekatan algoritma RF dengan variabel masukan yang berasal dari prediktor yang dihasilkan dari citra UAV untuk memprediksi massa tunas segar, jumlah buah tomat, dan massa hasil panen berdasarkan sejumlah variabel yang berasal dari citra UAV (Johansen dkk., 2020). Model RF digunakan untuk memprediksi hasil panen gandum, barley, dan canola dengan menggunakan tiga model terpisah yang dibuat berdasarkan kondisi pra-tanam, pertengahan musim, dan akhir musim untuk mengeksplorasi perubahan dalam kemampuan prediktif model seiring dengan bertambahnya informasi dalam musim (Filippi dkk., 2019). Algoritma *XGBoost Regressor* menunjukkan hasil yang jauh lebih baik dibandingkan dengan algoritma SVM, *Decision Tree*, dan RF dalam memprediksi hasil panen berdasarkan luas lahan pertanian, curah hujan, dan data suhu maksimum dan minimum (Ravi & Baranidharan, 2020). *Framework* prediksi hasil kedelai yang mengintegrasikan peningkatan gradien ekstrim (*XGBoost*) dan rekayasa fitur multidimensi dikembangkan di tingkat kabupaten di Amerika Serikat menggunakan dataset yang tersedia untuk umum (Li dkk., 2023). Model identifikasi area tanam gandum musim dingin berdasarkan algoritma *XGBoost* berhasil dibangun berdasarkan citra Sentinel-2, dengan mempertimbangkan 27 fitur spektral, tekstur, dan topograf, dimana algoritma *XCBoost* mengungguli tiga algoritme ML tradisional, yaitu RF, SVM dan CART (Wang dkk., 2023).

Penelitian ini memberikan kontribusi di bidang pertanian terkait pemilihan tanaman yang sesuai dengan kondisi lingkungan (suhu, kelembaban, pH, iklim) dan nutrisi tanah (N,P,K) untuk mendapatkan hasil panen yang optimal. Selain itu memberikan kontribusi terhadap penerapan *machine learning* dibidang pertanian dengan membandingkan dua algoritma, yaitu RF dan *XGBoost Regressor*.

## METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini seperti yang tersaji pada Gambar 1 terdiri dari beberapa tahap, yaitu: studi literature, pengumpulan data, pengolahan data awal, implementasi algoritma *machine learning*, pelatihan dan pengujian data, serta validasi dan evaluasi.



Sumber: Penelitian (2023)  
Gambar 1. Metode Penelitian

## 1. Studi Literatur

Pada tahap studi literatur dilakukan peninjauan literatur terkait pemilihan jenis tanaman berdasarkan faktor-faktor seperti iklim, tanah, kondisi lingkungan, dan nutrisi tanah. Analisis metode-metode penelitian sebelumnya yang digunakan dalam konteks penggunaan algoritma machine learning terhadap rekomendasi pemilihan jenis tanaman. Menemukan *state of the art* penelitian ini terhadap penelitian-penelitian terdahulu yang telah dilakukan terkait penerapan *machine learning* pada prediksi hasil panen pertanian dan rekomendasi tanaman yang sesuai dengan kondisi tanah dimana tanaman akan ditanam untuk mendapatkan hasil panen yang optimal.

## 2. Pengumpulan Data

Penelitian ini menggunakan kumpulan data (*dataset*) publik *crop recommendation* (Singh, 2023) yang merupakan kumpulan data yang memuat karakteristik tanah dan lingkungan yang optimal untuk pertumbuhan tanaman. Kumpulan data ini terdiri dari 8 atribut, yaitu: N, P, K, *temperature*, *humidity*, pH, *rainfall*, dan label (jenis tanaman) dengan jumlah banyaknya data sebesar 2200 baris data. Untuk jenis tanaman yang di rekomendasikan pada kumpulan data ini terdiri dari 22 jenis tanaman dengan masing-masing tanaman terdiri dari 100 data. Jenis tanaman dari dataset yang digunakan terdiri dari: *Rize*, *Maize*, *Jute*, *Cotton*, *Coconut*, *Papaya*, *Orange*, *Apple*, *Muskmelon*, *Watermelon*, *Grapes*, *Mango*, *Banana*, *Pomegranate*, *Lentil*, *Blackgram*, *Mungbean*, *Mothbeans*, *Pigeonpeas*, *Kidneybeans*, *Chickpea*, dan *Coffee*.

## 3. Pengolahan Data Awal

Pada tahap pengolahan data awal ini dilakukan proses pembersihan data dan normalisasi data dari kumpulan data. Pada kumpulan data *crop recommendation* yang digunakan pada penelitian ini tidak ditemukan data-data yang kosong (*null*). Pada atribut label tipe data yang digunakan dalam bentuk string atau karakter, sehingga untuk keperluan proses pemodelan menggunakan *machine learning*, data tersebut diubah ke dalam bentuk numerik. Pada pengolahan data awal ini dilakukan proses normalisasi data. Normalisasi fitur biasanya dilakukan untuk menyesuaikan dan mengnormalisasi nilai-nilai fitur, sehingga nilai-nilai fitur dapat dikurangi ke dalam rentang tertentu, seperti  $[-1.0, 1.0]$  atau  $[0, 1.0]$ . Tindakan normalisasi fitur membantu mengurangi perbedaan yang berlebihan dari berbagai rentang nilai fitur dan mengurangi ketergantungan algoritma pada unit pengukuran fitur. Umumnya, normalisasi fitur menggunakan transformasi linear, transformasi log, transformasi Tan, dan metode lain untuk standarisasi data, sehingga data dapat diubah ke dalam ruang bersama yang lebih kecil (Q. Xu & Yin, 2021). Pada Gambar 2, diperlihatkan kumpulan data sebelum dilakukan proses normalisasi, sedangkan pada Gambar 3 memperlihatkan hasil proses normalisasi kumpulan data *crop recommendation*.

	Nitrogen	phosphorus	potassium	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice
...	...	...	...	...	...	...	...	...
2195	107	34	32	26.774637	66.413269	6.780064	177.774507	coffee
2196	99	15	27	27.417112	56.636362	6.086922	127.924610	coffee
2197	118	33	30	24.131797	67.225123	6.362608	173.322839	coffee
2198	117	32	34	26.272418	52.127394	6.758793	127.175293	coffee
2199	104	18	30	23.603016	60.396475	6.779833	140.937041	coffee

2200 rows x 8 columns

Sumber: Penelitian (2023)  
Gambar 2. Kumpulan Data sebelum di Normalisasi

	Nitrogen	phosphorus	potassium	temperature	humidity	ph	rainfall	label_encoded
0	90	42	43	20.879744	82.002744	6.502985	202.935536	236.2
1	85	58	41	21.770462	80.319644	7.038096	226.655537	236.2
2	60	55	44	23.004459	82.320763	7.840207	263.964248	236.2
3	74	35	40	26.491096	80.158363	6.980401	242.864034	236.2
4	78	42	42	20.130175	81.604873	7.628473	262.717340	236.2
...	...	...	...	...	...	...	...	...
2195	107	34	32	26.774637	66.413269	6.780064	177.774507	158.1
2196	99	15	27	27.417112	56.636362	6.086922	127.924610	158.1
2197	118	33	30	24.131797	67.225123	6.362608	173.322839	158.1
2198	117	32	34	26.272418	52.127394	6.758793	127.175293	158.1
2199	104	18	30	23.603016	60.396475	6.779833	140.937041	158.1

2200 rows x 8 columns

Sumber: Penelitian (2023)  
Gambar 3. Kumpulan Data setelah di Normalisasi

#### 4. Implementasi Algoritma Machine Learning

Pada penelitian ini digunakan dibandingkan tingkat akurasi dua buah algoritma machine learning, yaitu: algoritma *Random Forest* dan algoritma *XGBoost Regressor* untuk merekomendasikan tanaman yang sesuai dengan kondisi lingkungan dan kadar nutrisi tanah.

Algoritma RF dibentuk dengan beberapa pohon keputusan, dan setiap pohon mendapatkan efek susunan posisinya dengan menggunakan klasifikasi yang berbeda. Pendekatan ini memungkinkan penilaian alokasi sampel dengan menggunakan teknik pengambilan sampel acak, yang juga sangat sesuai untuk model-model yang kecil (Murugan dkk., 2019).

#### 5. Pelatihan dan Pengujian Data

Pada *machine learning* peran penting data latih dan data uji tak terbantahkan dalam membangun model yang handal. Data latih sebagai dasar utama yang menyediakan bahan bagi model untuk mempelajari pola dan relasi antar variabel. Melalui interaksi dengan data latih, model dapat membentuk representasi internal yang

memungkinkannya membuat prediksi atau keputusan ketika dihadapkan pada data serupa. Sebaliknya, data uji, sebagai himpunan data yang terpisah, menjadi ukuran penting untuk mengevaluasi kinerja model. Dengan memanfaatkan kumpulan data ini dapat dilakukan pengujian seberapa baik model mampu beradaptasi dengan data yang belum pernah dilihatnya sebelumnya. Pemisahan yang jelas antara data latih dan data uji memastikan bahwa model tidak sekadar menghafal data pelatihan, melainkan mampu menghasilkan hasil yang akurat dan dapat diandalkan dalam berbagai situasi dunia nyata. Oleh karena itu, kebijakan penggunaan data latih dan data uji yang cerdas menjadi kunci utama dalam merancang model machine learning yang handal dan adaptif.

Pada penelitian ini, untuk membagi kumpulan data menjadi data latih dan data uji menggunakan *Splitting train and test set*. Proses pembentukan data latih dan data uji di perlihatkan pada Gambar 4. Pada proses Gambar 4, data latih ditentukan 80% dari kumpulan data yaitu sebanyak 1760 baris data dan data uji sebanyak 20% dari kumpulan data, yaitu sebanyak 440 baris data.

```
#Splitting train and test set
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=.2,random_state= 32)

X_train.shape
(1760, 7)

X_test.shape
(440, 7)

y_train.shape
(1760,)

y_test.shape
(440,)
```

Sumber: Penelitian (2023)

Gambar 4. Proses Pembentukan Data Latih dan Data Uji

## 6. Validasi dan Evaluasi

Untuk validasi dan evaluasi hasil prediksi rekomendasi tanaman menggunakan algoritma RF dan XGBoost digunakan tiga alat ukur, yaitu: Mean Absolute Error (MAE), Mean Square Error (MSE), dan  $R^2$  atau R-squared. Persamaan untuk MAE, MSE dan R-squared ditunjukkan pada persamaan (1), (2), dan (3) (Chicco dkk., 2021).

$$MAE = \frac{1}{m} \sum_{i=1}^m |X_i - Y_i| \quad (1)$$

$$MSE = \frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2 \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (\bar{Y} - Y_i)^2} ; \bar{Y} = \frac{1}{m} \sum_{i=1}^m Y_i \quad (3)$$

Keterangan:

$X_i$  = Nilai aktual

$Y_i$  = Nilai prediksi

$\bar{Y}$  = Nilai mean dari nilai sebenarnya

m = Jumlah sampel

i = Sampel ke-i

## HASIL DAN PEMBAHASAN

Untuk membandingkan tingkat akurasi menggunakan algoritma *random forest* dan algoritma *XGBoost Regressor* digunakan tiga alat ukur akurasi, yaitu MSE, MAE, dan  $R^2$ . Pada masing-masing algoritma dilakukan sepuluh kali pengujian tingkat akurasi. Pada Tabel 1, menunjukkan hasil pengujian menggunakan algoritma random Forest, hasilnya didapatkan rata-rata tingkat akurasi sebagai berikut: MSE didapatkan tingkat akurasi sebesar 36.73681574, MAE didapatkan hasil rata-rata sebesar 1.848396760, dan menggunakan  $R^2$  didapatkan hasil akurasi sebesar 0.98533501961927.

Tabel 1. Hasil Pengujian Algoritma *Random Forest*

Pengujian ke-	MSE	MAE	R <sup>2</sup>
1	36.18003079	1.866090909	0.98581883221623
2	34.71698025	1.821013636	0.98639229124164
3	33.89301159	1.806286363	0.98671525497282
4	38.22413768	1.867888636	0.98501762165197
5	38.55066330	1.939588636	0.98488963628514
6	37.88469704	1.826336363	0.98515066921088
7	39.74381712	1.895961363	0.98442196629931
8	37.29349415	1.846872727	0.98538239779370
9	35.47541704	1.796079545	0.98609501345656
10	35.40590844	1.770759090	0.98612225813076
<b>Rata-rata =</b>	<b>36.73681574</b>	<b>1.848396760</b>	<b>0.98533501961927</b>

Sumber: Penelitian (2023)

Pada Tabel 2, menunjukkan hasil pengukuran tingkat akurasi menggunakan MSE, MAE, dan R<sup>2</sup> untuk implementasi algoritma *XGBoost Regressor* untuk rekomendasi pemilihan jenis tanaman. Setelah dilakukan sepuluh kali pengujian didapatkan hasil yang sama untuk nilai akurasi MSE, MAE, dan R<sup>2</sup>. Hasil tingkat akurasi menggunakan algoritma *XGBoost Regressor* sebagai berikut: MSE didapatkan tingkat akurasi sebesar 37.17297890, MAE didapatkan hasil rata-rata sebesar 2.203242909, dan menggunakan R<sup>2</sup> didapatkan hasil akurasi sebesar 0.98542963509705.

Tabel 2. Hasil Pengujian Algoritma *XGBoost Regressor*

Pengujian ke-	MSE	MAE	R <sup>2</sup>
1	37.17297890	2.203242909	0.98542963509705
2	37.17297890	2.203242909	0.98542963509705
3	37.17297890	2.203242909	0.98542963509705
4	37.17297890	2.203242909	0.98542963509705
5	37.17297890	2.203242909	0.98542963509705
6	37.17297890	2.203242909	0.98542963509705
7	37.17297890	2.203242909	0.98542963509705
8	37.17297890	2.203242909	0.98542963509705
9	37.17297890	2.203242909	0.98542963509705
10	37.17297890	2.203242909	0.98542963509705
<b>Rata-rata =</b>	<b>37.17297890</b>	<b>2.203242909</b>	<b>0.98542963509705</b>

Sumber: Penelitian (2023)

Tabel 3. Perbandingan Akurasi Algoritma *Random Forest* dan *XGBoost Regressor*

Algoritma	MSE	MAE	R <sup>2</sup>
Random Forest	36.73681574	1.848396760	0.98533501961927
XGBoost Regressor	37.17297890	2.203242909	0.98542963509705

Sumber: Penelitian (2023)

Berdasarkan tiga nilai akurasi yang telah didapat dari hasil pengujian algoritma *Random Forest* (Tabel 1) dan *XGBoost* (Tabel 2) pada kumpulan data rekomendasi tanaman yang sesuai dengan kondisi lingkungan dan nutrisi tanah maka diperoleh hasil bahwa kedua algoritma mempunyai tingkat akurasi yang hampir sama seperti yang tersaji pada Tabel 3. Berdasarkan Tabel 3, untuk algoritma *Random Forest* tingkat hasil akurasinya cukup baik dibandingkan dengan algoritma *XGBoost Regressor* dengan menggunakan MSE dan MAE, namun dengan menggunakan R<sup>2</sup> atau *R-Square* diperoleh akurasi terbaik menggunakan algoritma *XGBoost Regressor*. Perbedaan tingkat akurasi yang didapatkan dalam penelitian ini tidak menunjukkan perbedaan yang cukup jauh, sehingga kedua algoritma ini dapat digunakan dengan baik untuk melakukan prediksi rekomendasi pemilihan

jenis tanaman pada lingkungan dan nutrisi tanah yang sesuai. Hasil pengujian prediksi rekomendasi tanaman berdasarkan kondisi lingkungan dan nutrisi media tanam menggunakan algoritma *Random Forest* dan *XGBoost Regressor* tersaji pada Tabel 4. Pada Tabel 4 ini, diberikan 5 kombinasi nilai-nilai N, P, K, Suhu, Humidity, pH, Rainfall dimasukkan pada model prediksi, sehingga didapatkan rekomendasi tanaman yang sesuai dengan nilai-nilai yang dimasukkan tersebut.

Tabel 4. Hasil Prediksi Rekomendasi Tanaman Algoritma *Random Forest* dan *XGBoost Regressor*

No	N	P	K	Suhu	Humidity	pH	Rainfall	Rekomendasi Tanaman
1	49	61	45	32.76795	94.57377	6.7642	240.479592	Papaya
2	3	136	205	17.58629	80.84806	6.3347	71.406545	Grapes
3	116	71	47	27.57278	82.06388	6.4357	91.342765	Banana
4	61	64	52	43.30204	92.83405	6.6411	110.562229	Papaya
5	25	65	21	33.86351	68.59232	6.8802	69.244641	Blackgram

Sumber: Penelitian (2023)

## KESIMPULAN

Penelitian ini telah melakukan investigasi tingkat akurasi dua buah algoritma Machine Learning, yaitu Random Forest dan XGBoost Regressor untuk melakukan prediksi rekomendasi tanaman yang sesuai dengan kondisi lingkungan dan nutrisi tanah. Hasil yang didapat menunjukkan bahwa kedua algoritma mempunyai tingkat akurasi yang sama baiknya. Algoritma Random Forest mempunyai tingkat akurasi yang baik diukur dengan menggunakan MSE dan MAE, yaitu sebesar 36.73681574 dan 1.848396760, sedangkan algoritma XGBoost Regressor mempunyai tingkat akurasi yang baik dengan menggunakan alat ukur akurasi  $R^2$  atau R-Square dengan tingkat akurasi sebesar 0.98542963509705. Hasil penelitian memberikan rekomendasi bahwa kedua algoritma yang digunakan pada penelitian ini dapat di implementasi untuk melakukan prediksi rekomendasi tanaman yang sesuai dengan kondisi lingkungan dan nutrisi tanah dimana tanaman tersebut akan ditanam.

## REFERENSI

- Banerjee, G., Sarkar, U., & Ghosh, I. (2021). *A Fuzzy Logic-Based Crop Recommendation System* (D. Bhattacharjee, D. K. Kole, N. Dey, S. Basu, & D. Plewczynski (ed.); hal. 57–69). Springer Singapore. [https://doi.org/10.1007/978-981-15-7834-2\\_6](https://doi.org/10.1007/978-981-15-7834-2_6)
- Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, 1–24. <https://doi.org/10.7717/PEERJ-CS.623>
- Craker, L. E. (2021). 4 Plant Growth Factors That Affect All Plants. *Journal of Soil Science & Plant Health*, 5(3).
- Filippi, P., Jones, E. J., Wimalathunge, N. S., Somarathna, P. D. S. N., Pozza, L. E., Ugbaje, S. U., Jephcott, T. G., Paterson, S. E., Whelan, B. M., & Bishop, T. F. A. (2019). An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. *Precision Agriculture*, 20(5), 1015–1029. <https://doi.org/10.1007/s11119-018-09628-4>
- Johansen, K., Morton, M. J. L., Malbeteau, Y., Aragon, B., Al-Mashharawi, S., Ziliani, M. G., Angel, Y., Fiene, G., Negrão, S., Mousa, M. A. A., Tester, M. A., & McCabe, M. F. (2020). Predicting Biomass and Yield in a Tomato Phenotyping Experiment Using UAV Imagery and Random Forest. *Frontiers in Artificial Intelligence*, 3(May). <https://doi.org/10.3389/frai.2020.00028>
- Li, Y., Zeng, H., Zhang, M., Wu, B., Zhao, Y., Yao, X., Cheng, T., Qin, X., & Wu, F. (2023). A county-level soybean yield prediction framework coupled with XGBoost and multidimensional feature engineering. *International Journal of Applied Earth Observation and Geoinformation*, 118, 103269. <https://doi.org/https://doi.org/10.1016/j.jag.2023.103269>
- Marques Ramos, A. P., Prado Osco, L., Elis Garcia Furuya, D., Nunes Gonçalves, W., Cordeiro Santana, D., Pereira Ribeiro Teodoro, L., Antonio da Silva Junior, C., Fernando Capristo-Silva, G., Li, J., Henrique Rojo Baio, F., Marcato Junior, J., Eduardo Teodoro, P., & Pistori, H. (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture*, 178, 105791. <https://doi.org/https://doi.org/10.1016/j.compag.2020.105791>
- Meshram, V., Patil, K., Meshram, V., Hanchate, D., & Ramkteke, S. D. (2021). Machine learning in agriculture domain: A state-of-art survey. *Artificial Intelligence in the Life Sciences*, 1, 100010.

- <https://doi.org/https://doi.org/10.1016/j.ailsi.2021.100010>
- Murugan, A., Nair, S. A. H., & Kumar, K. P. S. (2019). Detection of Skin Cancer Using SVM, Random Forest and kNN Classifiers. *Journal of Medical Systems*, 43(8), 269. <https://doi.org/10.1007/s10916-019-1400-8>
- Pande, S M, Ramesh, P. K., Anmol, A., Aishwarya, B. R., Rohoilla, K., & Shaurya, K. (2021). Crop Recommender System Using Machine Learning Approach. *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 1066–1071. <https://doi.org/10.1109/ICCMC51019.2021.9418351>
- Pande, Shilpa Mangesh, Ramesh, P. K., Anmol, A., Aishwarya, B. R., Rohilla, K., & Shaurya, K. (2021). Crop Recommender System Using Machine Learning Approach. *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021, Iccmc*, 1066–1071. <https://doi.org/10.1109/ICCMC51019.2021.9418351>
- Ravi, R., & Baranidharan, B. (2020). Crop Yield Prediction using XG Boost Algorithm. *International Journal of Recent Technology and Engineering (IJRTE)*, 8(5), 3516–3520. <https://doi.org/10.35940/ijrte.d9547.018520>
- Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2021). Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access*, 9, 4843–4873. <https://doi.org/10.1109/ACCESS.2020.3048415>
- Singh, R. (2023). *Crop recommendation data* (V1 ed.). Harvard Dataverse. <https://doi.org/doi:10.7910/DVN/4GBWFV>
- Supriya, B. N., & Akki, C. B. (2021). Sentiment prediction using enhanced xgboost and tailored random forest. *International Journal of Computing and Digital Systems*, 10(1), 191–199. <https://doi.org/10.12785/ijcds/100119>
- Upadhyay, H., Juneja, A., Turabieh, H., Malik, S., Gupta, A., Bitsue, Z. K., & Upadhyay, C. (2022). Exploration of Crucial Factors Involved in Plants Development Using the Fuzzy AHP Method. *Mathematical Problems in Engineering*, 2022. <https://doi.org/10.1155/2022/4279694>
- Wang, Y., Zhu, D., & Ding, Y. (2023). Identification of Winter Wheat- Identification of Winter Wheat-Growing Areas Based on the XGBoost Algorithm. *Preprints*. <https://doi.org/10.20944/preprints202303.0346.v1>
- Xu, Q., & Yin, J. (2021). Application of Random Forest Algorithm in Physical Education. *Scientific Programming*, 2021, 1996904. <https://doi.org/10.1155/2021/1996904>
- Xu, X., Gao, P., Zhu, X., Guo, W., Ding, J., Li, C., Zhu, M., & Wu, X. (2019). Design of an integrated climatic assessment indicator (ICAI) for wheat production: A case study in Jiangsu Province, China. *Ecological Indicators*, 101, 943–953. <https://doi.org/https://doi.org/10.1016/j.ecolind.2019.01.059>