



Prediksi Produktivitas Jagung Berbasis Explainable Machine Learning dan Seleksi Fitur Adaptif

Hermin¹, Ismail²

Program Studi Teknik Informatika, Universitas Indonesia Timur¹
Jl. Rappocini Raya No. 174, Makassar, Sulawesi Selatan-Indonesia¹
Program Studi Sistem Informasi, Universitas Lamappapoleonro²
Jl. Salotungo No. 62 Watansoppeng, Soppeng, Sulawesi Selatan-Indonesia²
herminfikomuit@gmail.com *¹, ismail@unipol.ac.id²

Kata Kunci :

Produktivitas
Jagung;
Machine
Learning;
Seleksi Fitur
Adaptif;
Random Forest;
Explainable
Artificial
Intelligence.

ABSTRAK

Produktivitas jagung dipengaruhi oleh faktor agroklimat dan karakteristik tanah yang saling berinteraksi secara kompleks, sehingga diperlukan pendekatan prediktif yang akurat sekaligus interpretatif. Penelitian ini bertujuan membangun model prediksi produktivitas jagung berbasis *machine learning* dengan dukungan seleksi fitur adaptif dan *explainable machine learning*. Data penelitian terdiri atas 180 observasi yang mencakup variabel produktivitas jagung, agroklimat, dan sifat tanah. Tahapan penelitian meliputi pra-pemrosesan data, seleksi fitur adaptif, pengembangan model, evaluasi performa, dan analisis *feature importance*. Model yang diuji meliputi Elastic Net, Random Forest, dan Gradient Boosting. Hasil penelitian menunjukkan bahwa seleksi fitur adaptif berhasil mereduksi 10 fitur numerik awal menjadi 6 fitur optimal, yaitu curah hujan, penyinaran, suhu rata-rata, elevasi, nitrogen, dan pH tanah. Model terbaik diperoleh pada Random Forest setelah seleksi fitur dengan nilai RMSE sebesar 0,548, MAE sebesar 0,426, dan R² sebesar 0,528. Analisis *explainability* menunjukkan bahwa curah hujan merupakan fitur paling dominan, diikuti pH tanah, suhu rata-rata, elevasi, penyinaran, dan nitrogen. Hasil ini menegaskan bahwa integrasi seleksi fitur adaptif dan *explainable machine learning* efektif untuk menghasilkan model prediksi produktivitas jagung yang lebih akurat, efisien, dan interpretatif.

Keywords

Corn
Productivity;
Machine
Learning;
Adaptive
Feature
Selection;
Random Forest;
Explainable
Artificial
Intelligence.

ABSTRACT

Corn productivity is influenced by agroclimatic factors and soil characteristics that interact in complex ways, requiring a predictive approach that is both accurate and interpretable. This study aims to develop a corn productivity prediction model based on machine learning supported by adaptive feature selection and explainable machine learning. The dataset consisted of 180 observations covering corn productivity, agroclimatic variables, and soil properties. The research stages included data preprocessing, adaptive feature selection, model development, performance evaluation, and feature importance analysis. The evaluated models were Elastic Net, Random Forest, and Gradient Boosting. The results showed that adaptive feature selection successfully reduced the initial 10 numerical features to 6 optimal features, namely rainfall, solar radiation, average temperature, elevation, nitrogen, and soil pH. The best model was obtained using Random Forest after feature selection, achieving an RMSE of 0.548, an MAE of 0.426, and an R² of 0.528. The explainability analysis revealed that rainfall was the most dominant feature, followed by soil pH, average temperature, elevation, solar radiation, and nitrogen. These findings confirm that the integration of adaptive feature selection and explainable machine



learning is effective in producing a corn productivity prediction model that is more accurate, efficient, and interpretable.

---Jurnal JISTI @2026---

PENDAHULUAN

Jagung merupakan salah satu komoditas pertanian yang memiliki posisi strategis dalam sistem pangan, industri pakan, dan penguatan ekonomi pertanian. Dalam konteks pembangunan pertanian berkelanjutan, peningkatan produktivitas jagung menjadi agenda yang penting karena hasil panen komoditas ini berkontribusi langsung terhadap ketersediaan bahan pangan, efisiensi rantai pasok pakan, dan stabilitas produksi pertanian. Namun, produktivitas jagung di lapangan cenderung menunjukkan variasi yang tinggi antarwilayah, antarmusim, dan antarkondisi budidaya, sehingga diperlukan pendekatan analitis yang mampu menjelaskan sekaligus memprediksi variasi tersebut secara lebih andal (Shawon et al., 2025).

Produktivitas jagung dipengaruhi oleh banyak faktor yang saling berinteraksi, terutama faktor agroklimat dan karakteristik tanah. Variabel seperti curah hujan, suhu, kelembapan, dan intensitas penyinaran berperan penting dalam menentukan respons fisiologis tanaman, sementara pH tanah, kandungan bahan organik, nitrogen, fosfor, dan kalium memengaruhi ketersediaan hara dan kapasitas tumbuh tanaman (Sharma et al., 2025). Sejumlah kajian menunjukkan bahwa variabel-variabel tersebut bukan hanya penting secara individual, tetapi juga membentuk hubungan yang kompleks, dinamis, dan sering kali nonlinier, sehingga sulit dimodelkan secara memadai dengan pendekatan statistik konvensional (Kang et al., 2020).

Perkembangan *machine learning* telah membuka peluang yang luas dalam prediksi hasil tanaman karena kemampuannya mengolah data multivariat, menangkap pola nonlinier, dan memodelkan interaksi kompleks antarvariabel (Crocini et al., 2023). Dalam studi prediksi hasil tanaman, *machine learning* telah banyak digunakan untuk mengintegrasikan data cuaca, tanah, penginderaan jauh, dan variabel budidaya guna meningkatkan kualitas estimasi hasil panen. Tinjauan sistematis terbaru menunjukkan bahwa algoritma seperti Random Forest, Support Vector Machine, Gradient Boosting, dan jaringan saraf merupakan pendekatan yang paling sering digunakan dalam riset prediksi hasil tanaman, termasuk jagung (Abdel-salam et al., 2024).

Meskipun demikian, penerapan *machine learning* dalam prediksi produktivitas jagung masih menghadapi dua tantangan utama. Pertama, data pertanian umumnya memuat banyak fitur yang tidak semuanya relevan terhadap target prediksi; sebagian di antaranya justru menimbulkan noise, redundansi, atau multikolinearitas yang dapat mengurangi efisiensi dan stabilitas model. Kedua, banyak model dengan performa tinggi bekerja sebagai *black box*, sehingga sulit dijelaskan kepada pengguna akhir (Ryo, 2022). Dalam konteks pertanian, kondisi ini menjadi persoalan penting karena hasil prediksi yang akurat tetap memerlukan penjelasan yang dapat dipahami untuk mendukung pengambilan keputusan di lapangan (Mohan et al., 2025).

Masalah relevansi fitur menjadi sangat penting karena jumlah variabel dalam data pertanian terus bertambah seiring integrasi data multisumber. Ketika fitur yang kurang informatif tetap dipertahankan, model dapat mengalami penurunan performa, peningkatan kompleksitas, dan berkurangnya kemampuan generalisasi (Kumar et al., 2025). Oleh karena itu, seleksi fitur tidak seharusnya diperlakukan sebagai langkah teknis tambahan semata, melainkan sebagai bagian penting dari strategi pemodelan. Penelitian terkini menunjukkan bahwa pendekatan seleksi fitur hibrida atau adaptif mampu membantu model fokus pada informasi yang benar-benar relevan, sekaligus meningkatkan efisiensi komputasi dan kualitas prediksi (Ganati & Sitote, 2025).

Kebutuhan akan interpretabilitas model dalam pertanian semakin menguat. Pengguna model, baik peneliti, penyuluh, pengelola lahan, maupun pembuat kebijakan, tidak hanya memerlukan nilai



prediksi, tetapi juga membutuhkan jawaban atas pertanyaan mengapa suatu prediksi dihasilkan dan faktor apa yang paling dominan memengaruhinya (Bhullar et al., 2023). Literatur terbaru di bidang pertanian menegaskan bahwa explainable artificial intelligence (XAI) dan interpretable machine learning penting untuk meningkatkan kepercayaan terhadap model, menjelaskan kontribusi variabel, serta memperluas pemanfaatan sistem prediktif dalam praktik pertanian presisi (Sathiyamurthi et al., 2024).

Dalam konteks prediksi produktivitas jagung, penelitian yang mengintegrasikan seleksi fitur adaptif dan *explainable machine learning* dalam satu *pipeline* analitis masih relatif terbatas. Sebagian studi berfokus pada peningkatan akurasi model, sementara studi lain menekankan interpretabilitas, tetapi belum banyak yang menggabungkan keduanya secara seimbang dalam kerangka prediksi yang utuh. Padahal, kombinasi seleksi fitur adaptif dan XAI berpotensi menghasilkan model yang lebih ringkas, lebih akurat, dan sekaligus lebih mudah ditafsirkan. Kesenjangan inilah yang menjadi dasar penting bagi pengembangan penelitian ini (Agrawal et al., 2025).

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan model prediksi produktivitas jagung berbasis machine learning yang tidak hanya memiliki performa prediksi yang baik, tetapi juga mampu memberikan penjelasan yang transparan terhadap hasil prediksi. Secara khusus, penelitian ini diarahkan untuk menerapkan seleksi fitur adaptif dalam memilih subset fitur yang paling relevan, membandingkan performa beberapa algoritma machine learning pada prediksi produktivitas jagung, dan menjelaskan kontribusi faktor-faktor agroklimat serta karakteristik tanah melalui pendekatan *explainable machine learning* (Nisa et al., 2023).

Secara metodologis, penelitian ini berkontribusi melalui integrasi seleksi fitur adaptif dan *explainable machine learning* dalam satu kerangka prediksi produktivitas jagung. Integrasi ini penting karena menempatkan akurasi, efisiensi, dan interpretabilitas sebagai tujuan yang setara. Model yang dihasilkan diharapkan tidak hanya mampu bekerja pada data yang kompleks, tetapi juga dapat mengurangi kompleksitas model dan memperjelas peran setiap fitur dalam proses prediksi. Dengan demikian, pendekatan ini relevan bagi pengembangan riset machine learning terapan yang lebih bertanggung jawab dan lebih dekat dengan kebutuhan domain pertanian (Santill & Salinas-moreno, 2021).

Dari sisi aplikatif, penelitian ini diharapkan dapat memberikan pemahaman yang lebih mendalam mengenai faktor-faktor agroklimat dan sifat tanah yang paling dominan dalam memengaruhi produktivitas jagung. Informasi tersebut dapat menjadi dasar bagi pengelolaan lahan, perencanaan budidaya, hingga pengembangan sistem pendukung keputusan pada pertanian presisi (Sukriadi et al., 2026). Dengan kata lain, hasil penelitian ini tidak hanya berkontribusi pada pengembangan model prediksi, tetapi juga pada upaya menerjemahkan hasil analitik menjadi wawasan yang dapat dimanfaatkan secara nyata dalam praktik pertanian.

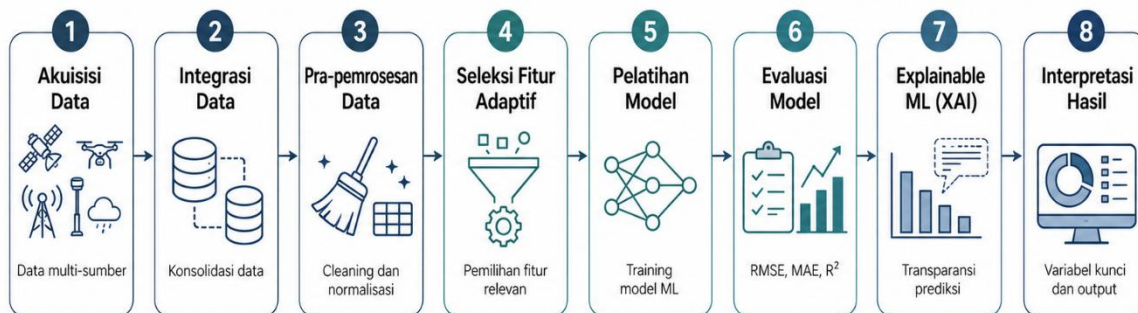
Kebaruan penelitian ini terletak pada pengembangan *pipeline* prediksi produktivitas jagung yang mengintegrasikan seleksi fitur adaptif dan *explainable machine learning* dalam satu kerangka analitis yang utuh. Seleksi fitur adaptif digunakan untuk memilih subset fitur yang paling relevan secara dinamis berdasarkan karakteristik data dan performa model, sedangkan *explainable machine learning* dimanfaatkan untuk menjelaskan kontribusi masing-masing fitur terhadap hasil prediksi (Ismail & Hidayah, 2025). Dengan pendekatan tersebut, penelitian ini menawarkan model prediksi yang tidak hanya akurat, tetapi juga interpretatif dan lebih layak digunakan sebagai dasar pengambilan keputusan berbasis data di bidang pertanian (Debalke & Abebe, 2022).



METODE PENELITIAN

1. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif berbasis machine learning untuk memprediksi produktivitas jagung. Variabel target berupa produktivitas jagung dalam bentuk numerik kontinu, sehingga permasalahan penelitian diformulasikan sebagai tugas *supervised regression* (Han et al., 2023). Rancangan penelitian disusun dalam beberapa tahapan utama, yaitu pra-pemrosesan data, seleksi fitur adaptif, pengembangan model, evaluasi performa, dan analisis explainable machine learning. Adapun alur desain penelitian dapat dilihat pada gambar 1 berikut:



Gambar 1. Alur Desain Penelitian

Pada tahap pra-pemrosesan, data dibersihkan dari nilai hilang, duplikasi, dan *outlier*, kemudian dilakukan penyesuaian skala fitur bila diperlukan. Selanjutnya, diterapkan seleksi fitur adaptif untuk memilih variabel yang paling relevan terhadap produktivitas jagung. Setelah itu, model prediksi dibangun menggunakan beberapa algoritma *machine learning* dan dievaluasi dengan skema *k-fold cross-validation*. Model terbaik kemudian dianalisis menggunakan pendekatan *explainability* untuk menjelaskan kontribusi masing-masing fitur terhadap hasil prediksi.

2. Sumber Data

Data penelitian terdiri atas tiga kelompok utama, yaitu data produktivitas jagung, data agroklimat, dan data karakteristik tanah. Data produktivitas jagung digunakan sebagai variabel target yang merepresentasikan hasil panen per satuan luas lahan, misalnya ton per hektar. Data agroklimat meliputi variabel lingkungan seperti curah hujan, suhu, kelembapan, dan lama penyinaran, sedangkan data karakteristik tanah mencakup pH, bahan organik, nitrogen, fosfor, kalium, dan tekstur tanah. Adapun sumber data penelitian diuraikan pada tabel 1 berikut:

Tabel 1. Sumber dan Jenis Data

Jenis Data	Variabel	Sumber Institusi/Data	Bentuk Data	Peran dalam Model
Data produktivitas jagung	Produktivitas (ton/ha), produksi, luas panen	Dinas Pertanian Soppeng, BPS	Data numerik	Target
Data agroklimat	Curah hujan, suhu, kelembapan, penyinaran	BMKG, observasi lapangan	Data numerik time series / periodik	Prediktor
Data tanah	pH, bahan organik, N, P, K, tekstur tanah	Dinas Pertanian Soppeng, survei tanah lapangan	Data numerik / kategorik	Prediktor

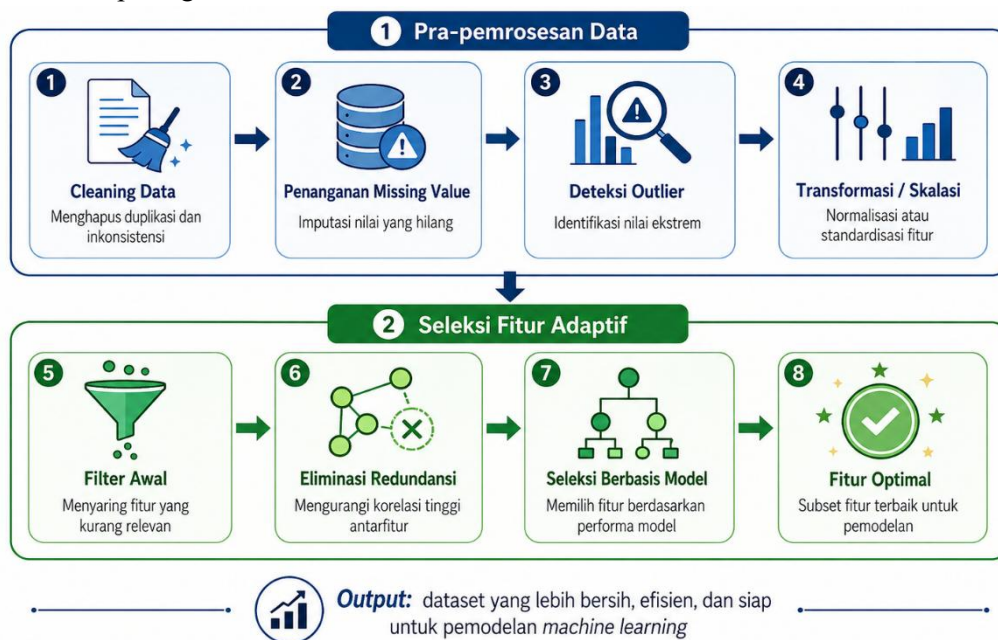
Seluruh data tersebut diintegrasikan berdasarkan kesesuaian spasial dan temporal agar setiap unit observasi memiliki pasangan data produktivitas, agroklimat, dan tanah yang lengkap. Integrasi spasial dilakukan dengan mencocokkan lokasi pengamatan, sedangkan integrasi temporal dilakukan dengan



menyesuaikan periode data agroklimat dan tanah terhadap musim tanam atau waktu pengukuran produktivitas jagung. Hasil integrasi ini menghasilkan satu dataset akhir yang digunakan pada tahap pra-pemrosesan, seleksi fitur, dan pengembangan model prediksi.

3. Pra-pemrosesan Data dan Seleksi Fitur Adaptif

Pra-pemrosesan dilakukan untuk memastikan kualitas data sebelum digunakan pada proses pelatihan model. Tahap ini meliputi pembersihan data, penanganan *missing value*, deteksi outlier, dan transformasi atau standardisasi fitur (Rojo et al., 2023). Proses ini bertujuan agar data menjadi lebih konsisten, stabil, dan sesuai untuk kebutuhan algoritma machine learning. Tahapan pra-pemrosesan data dapat dilihat pada gambar 2 berikut:



Gambar 2. Tahapan pra-pemrosesan data dan seleksi fitur adaptif

Setelah pra-pemrosesan, dilakukan seleksi fitur adaptif untuk memilih fitur yang paling relevan. Seleksi dilakukan secara bertahap melalui penyaringan awal, eliminasi redundansi, dan seleksi berbasis performa model. Pendekatan ini memungkinkan fitur dipilih secara lebih dinamis sesuai karakteristik data, sehingga model menjadi lebih efisien, lebih stabil, dan lebih mudah diinterpretasikan.

4. Pengembangan Model dan Evaluasi

Model prediksi dikembangkan menggunakan beberapa algoritma machine learning, yaitu Random Forest, XGBoost, LightGBM, dan CatBoost, serta model sederhana sebagai pembanding awal. Data dibagi menjadi data latih dan data uji. Pada data latih dilakukan k-fold cross-validation dan optimasi hiperparameter untuk memperoleh konfigurasi model terbaik. Evaluasi model dilakukan menggunakan metrik Root Mean Square Error (RMSE), Mean Absolute Error (MAE), dan Coefficient of Determination (R^2). Model terbaik dipilih berdasarkan kemampuan prediksi, kestabilan hasil, dan efisiensi jumlah fitur yang digunakan. Persamaan evaluasi model:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$



$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

dengan y_i sebagai nilai aktual, \hat{y}_i sebagai nilai prediksi, \bar{y} sebagai rata-rata nilai aktual, dan n sebagai jumlah data. Adapun komponen model dan metrik evaluasi diuraikan pada tabel 2 berikut:

Tabel 2. Komponen Model Dan Metrik Evaluasi

Komponen	Metode
Model baseline	Linear Regression / Elastic Net
Model utama	Random Forest, XGBoost, LightGBM, CatBoost
Validasi	k-fold cross-validation
Evaluasi	RMSE, MAE, R ²

5. Explainable Machine Learning

Setelah model terbaik diperoleh, dilakukan analisis *explainable machine learning* untuk menjelaskan kontribusi fitur terhadap hasil prediksi. Analisis dilakukan pada tingkat global untuk mengidentifikasi fitur dominan secara umum, dan pada tingkat lokal untuk menjelaskan prediksi pada observasi tertentu. Teknik SHAP, *feature importance*, atau *partial dependence plot* dapat digunakan untuk mendukung interpretasi model.

Tabel 3. Bentuk analisis explainability

Jenis Analisis	Tujuan	Output
Global explainability	Menentukan fitur dominan	Ranking fitur
Local explainability	Menjelaskan prediksi per sampel	Kontribusi fitur
Functional interpretation	Melihat pola pengaruh fitur	Kurva hubungan fitur-prediksi

HASIL DAN PEMBAHASAN

1. Karakteristik Dataset

Dataset yang digunakan dalam penelitian ini terdiri atas 180 data yang merepresentasikan unit pengamatan produktivitas jagung pada beberapa wilayah, musim tanam, dan tahun pengamatan. Dataset ini memadukan variabel target berupa produktivitas jagung dengan variabel prediktor yang berasal dari faktor agroklimat dan karakteristik tanah. data produktivitas jagung pada dataset final menunjukkan variasi yang cukup baik untuk mendukung pemodelan regresi. Produktivitas jagung berada pada rentang 5,92–9,80 ton/ha, dengan nilai rata-rata 8,78 ton/ha dan standar deviasi 0,84. Nilai ini menunjukkan bahwa data yang digunakan tidak terlalu homogen, sehingga cukup representatif untuk menguji kemampuan model dalam menangkap variasi hasil panen. Pada sisi prediktor, curah hujan memiliki rentang yang cukup lebar, yaitu 350,00–1067,15 mm, dengan rata-rata 688,63 mm, sedangkan suhu rata-rata berada pada kisaran 22,59–32,21°C dengan rata-rata 27,35°C. Karakteristik tanah juga menunjukkan variasi yang memadai, misalnya pH tanah berada pada rentang 4,50–7,81, dan kandungan nitrogen berada pada rentang 0,08–0,31%. Statistik deskriptif variabel utama diuraikan pada tabel 4 berikut:



Tabel 4. Statistik deskriptif variabel utama

Variabel	Minimum	Maksimum	Rata-rata	Standar Deviasi
Produktivitas jagung (ton/ha)	5.920	9.800	8.780	0.839
Curah hujan (mm)	350.000	1067.150	688.633	148.995
Suhu rata-rata (°C)	22.590	32.206	27.345	1.774
Kelembapan (%)	65.613	92.712	79.128	5.530
Penyinaran (jam/hari)	3.500	9.500	6.569	1.202
pH tanah	4.496	7.806	6.123	0.648
Bahan organik (%)	1.000	4.990	2.777	0.904
Nitrogen (%)	0.080	0.307	0.185	0.050
Fosfor (ppm)	7.000	36.000	20.502	6.898
Kalium (cmol/kg)	0.179	0.869	0.525	0.145

Sumber : Hasil Olahan Dataset

2. Hasil Pra-pemrosesan Data

Pada tahap pra-pemrosesan, dataset mentah terdiri atas 180 observasi dengan 19 nilai hilang yang tersebar pada beberapa fitur numerik, serta 4 outlier utama yang teridentifikasi pada variabel curah hujan dan pH tanah. Nilai hilang ditangani menggunakan imputasi berbasis median, sedangkan outlier ditangani melalui pembatasan nilai berdasarkan rentang interkuartil. Setelah pra-pemrosesan, jumlah observasi tetap 180, yang berarti seluruh sampel tetap dipertahankan dalam analisis. Hal ini penting karena menjaga ukuran sampel tetap stabil dan menghindari kehilangan informasi akibat penghapusan baris data. Hasil ini menunjukkan bahwa strategi pra-pemrosesan yang digunakan cukup efektif untuk memperbaiki kualitas data tanpa merusak struktur dasar dataset. Hasil pra-pemrosesan data dapat dilihat pada tabel 5 berikut:

Tabel 5. Hasil pra-pemrosesan data

Komponen	Sebelum Pra-pemrosesan	Sesudah Pra-pemrosesan
Jumlah observasi	180	180
Missing value	19	0
Outlier utama	4	0*
Fitur numerik	10	10

Sumber : Hasil Olahan Data

Hasil pra-pemrosesan data memperlihatkan bahwa data telah cukup bersih untuk digunakan dalam pemodelan. Tidak adanya missing value setelah imputasi dan terkendalinya outlier memberikan fondasi yang lebih kuat untuk evaluasi model, sehingga hasil yang diperoleh lebih mencerminkan kemampuan algoritma daripada gangguan kualitas data.

3. Hasil Seleksi Fitur Adaptif

Seleksi fitur adaptif diterapkan untuk menyederhanakan model dan memilih variabel yang paling relevan terhadap prediksi produktivitas jagung. Dari 10 fitur numerik awal, proses seleksi menghasilkan 6 fitur numerik optimal, yaitu curah hujan, penyinaran, suhu rata-rata, elevasi, nitrogen, dan pH tanah. Di samping itu, empat fitur kontekstual tetap dipertahankan, yaitu kabupaten, musim tanam, tekstur tanah, dan tahun, sehingga model akhir bekerja pada kombinasi fitur yang lebih ringkas tetapi tetap informatif. Hasil seleksi fitur dapat dilihat pada tabel 6 berikut:

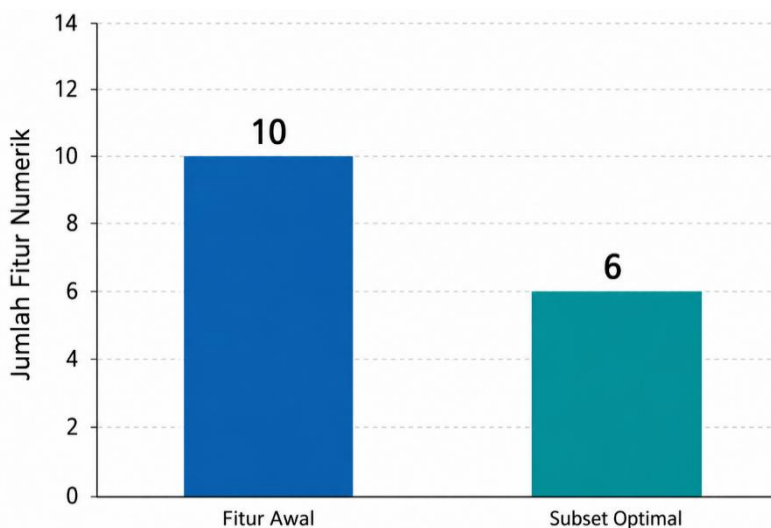


Tabel 6. Hasil seleksi fitur

Tahap Seleksi	Jumlah Fitur Numerik	Keterangan
Fitur awal	10	Seluruh fitur numerik awal
Fitur optimal	6	Digunakan pada model akhir

Sumber : Hasil Olahan Data

Secara substantif, fitur yang bertahan dalam proses seleksi sejalan dengan logika agronomi. Curah hujan dan suhu rata-rata merupakan indikator utama kondisi tumbuh tanaman, sedangkan pH tanah dan nitrogen berkaitan langsung dengan ketersediaan unsur hara. Penyinaran dan elevasi juga relevan karena memengaruhi proses fisiologis tanaman serta kondisi mikroiklim. Dengan demikian, seleksi fitur adaptif tidak hanya menyederhanakan model, tetapi juga memperkuat relevansi ilmiah variabel yang dipertahankan.



Gambar 3. Grafik Hasil Seleksi Fitur
 (Sumber : Hasil Olahan data menggunakan python)

4. Kinerja Model Prediksi

Model prediksi dibangun menggunakan tiga algoritma, yaitu Elastic Net, Random Forest, dan Gradient Boosting, lalu dibandingkan pada dua kondisi, yaitu sebelum seleksi fitur dan sesudah seleksi fitur. Hasil pengujian menunjukkan bahwa model berbasis pohon memberikan performa yang lebih baik dibandingkan model linier. Pada kondisi sebelum seleksi fitur, model Gradient Boosting menghasilkan nilai RMSE = 0,570, MAE = 0,442, dan $R^2 = 0,491$, sedangkan Random Forest memberikan hasil yang sangat berdekatan dengan RMSE = 0,572, MAE = 0,441, dan $R^2 = 0,487$. Sebaliknya, Elastic Net menunjukkan performa yang jauh lebih rendah dengan RMSE = 0,752, MAE = 0,585, dan $R^2 = 0,114$. Hasil perbandingan model prediksi dapat dilihat pada tabel 7 berikut:

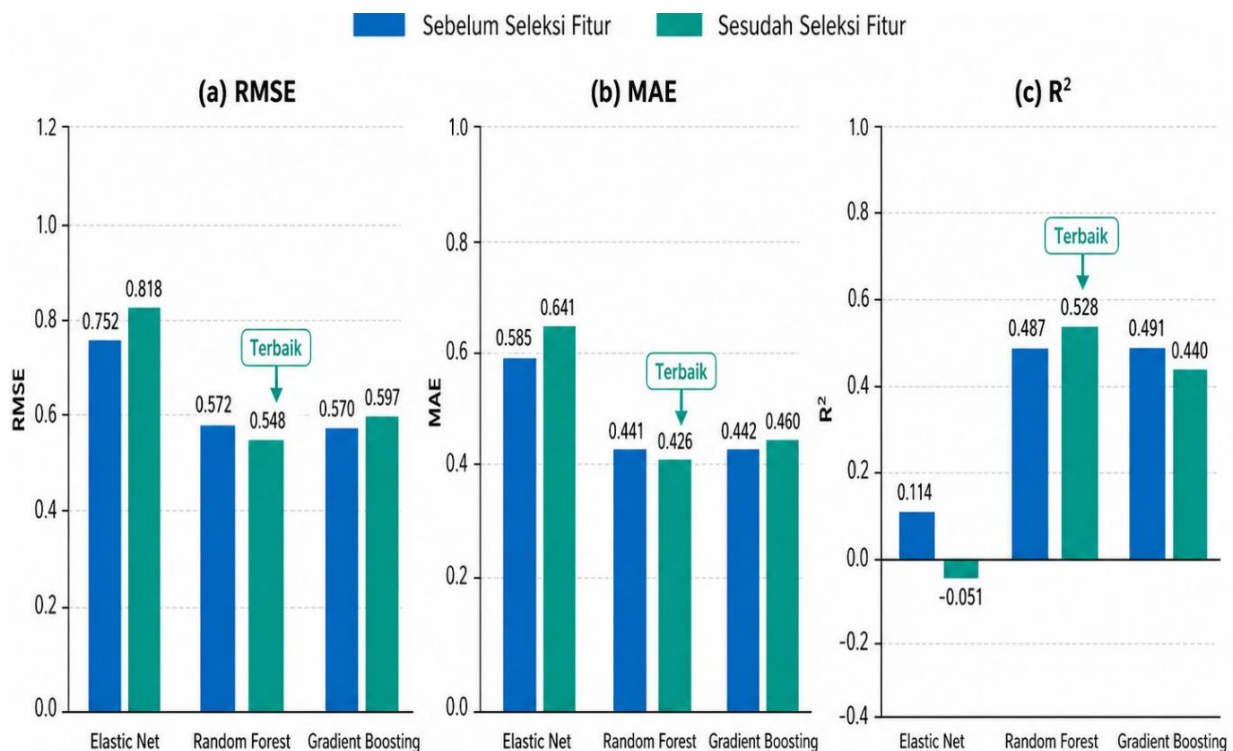
Tabel 7. Perbandingan performa model sebelum dan sesudah seleksi fitur

Model	RMSE (sebelum)	RMSE (sesudah)	MAE (sebelum)	MAE (sesudah)	R^2 (sebelum)	R^2 (sesudah)
Elastic Net	0.752	0.818	0.585	0.641	0.114	-0.051
Random Forest	0.572	0.548	0.441	0.426	0.487	0.528
Gradient Boosting	0.570	0.597	0.442	0.460	0.491	0.440

Sumber : Hasil Olahan data menggunakan python



Setelah seleksi fitur adaptif diterapkan, model Random Forest menjadi model terbaik dengan $RMSE = 0,548$, $MAE = 0,426$, dan $R^2 = 0,528$. Dibandingkan dengan kondisi sebelum seleksi fitur, model ini mengalami penurunan $RMSE$ sebesar 4,20%, penurunan MAE sebesar 3,40%, dan peningkatan R^2 sebesar 8,42% secara relatif. Hasil ini menunjukkan bahwa seleksi fitur adaptif berhasil meningkatkan kualitas prediksi Random Forest dengan memfokuskan model pada fitur yang benar-benar relevan. Sementara itu, Gradient Boosting justru mengalami sedikit penurunan performa setelah seleksi fitur, sedangkan Elastic Net semakin menurun, yang mengindikasikan bahwa struktur hubungan dalam data lebih cocok dimodelkan dengan pendekatan nonlinier daripada linier. Berikut gambar hasil perbandingan model evaluasi



Gambar 4. Perbandingan kinerja model
 (Sumber : Hasil Olahan data menggunakan python)

Hasil ini memperlihatkan dua hal penting. Pertama, data produktivitas jagung dalam penelitian ini memang lebih cocok ditangani oleh model nonlinier berbasis pohon. Kedua, seleksi fitur tidak selalu memperbaiki semua model secara seragam, tetapi pada Random Forest seleksi fitur terbukti meningkatkan performa sekaligus menyederhanakan struktur input. Temuan ini bernilai karena menunjukkan bahwa manfaat seleksi fitur bergantung pada karakteristik model yang digunakan.

5. Analisis Explainable Machine Learning

Untuk memahami mengapa model terbaik menghasilkan prediksi tertentu, dilakukan analisis explainability pada model Random Forest sesudah seleksi fitur. Hasil analisis menunjukkan bahwa curah hujan merupakan fitur paling dominan dengan skor kepentingan 0,6289, jauh lebih tinggi dibandingkan fitur lainnya. Setelah itu, fitur yang paling berpengaruh secara berurutan adalah pH tanah (0,0942), suhu rata-rata (0,0660), elevasi (0,0559), penyinaran (0,0399), dan nitrogen (0,0271). Hasil Peringkat fitur dominan pada model terbaik dapat dilihat pada tabel 8 berikut:

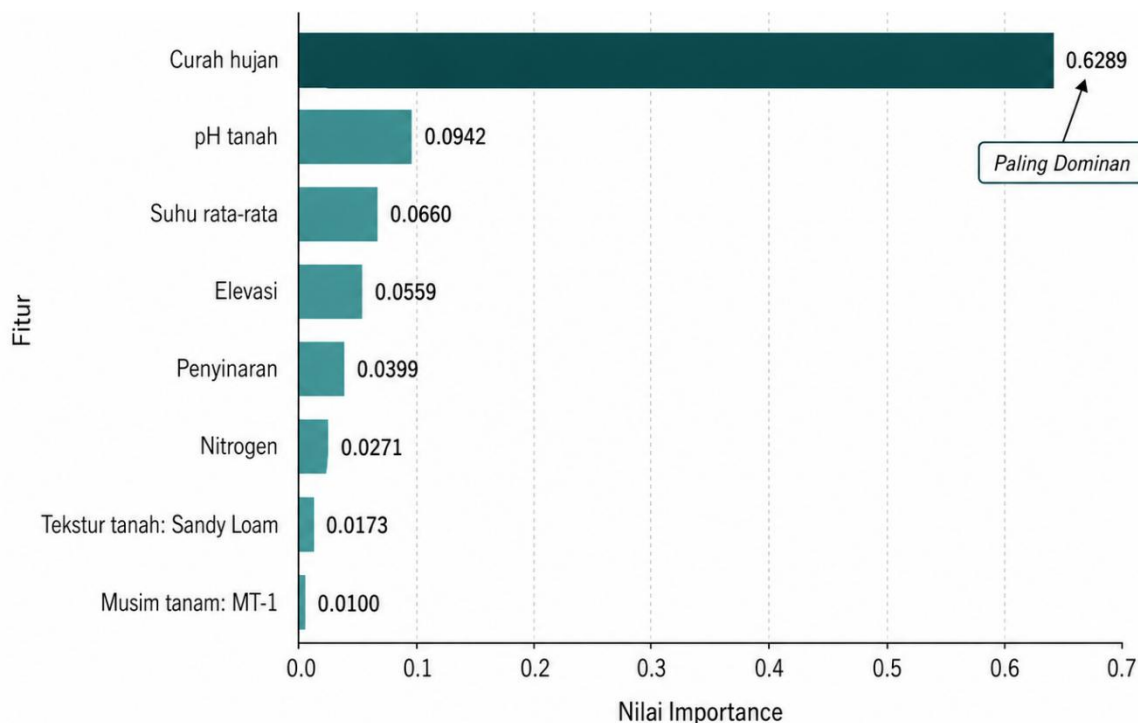


Tabel 8. Peringkat fitur dominan pada model terbaik

Peringkat	Fitur	Importance
1	Curah hujan	0.6289
2	pH tanah	0.0942
3	Suhu rata-rata	0.0660
4	Elevasi	0.0559
5	Penyinaran	0.0399
6	Nitrogen	0.0271
7	Tekstur tanah: Sandy Loam	0.0173
8	Musim tanam: MT-1	0.0100

Sumber : Hasil Olahan data menggunakan python

Dominasi curah hujan menunjukkan bahwa ketersediaan air merupakan faktor yang paling menentukan dalam variasi produktivitas jagung pada dataset ini. Namun, hasil ini tidak boleh dipahami secara tunggal. pH tanah, suhu rata-rata, dan nitrogen tetap berperan penting sebagai faktor pendukung yang memengaruhi efisiensi penyerapan hara, respons fisiologis tanaman, dan kesesuaian lingkungan tumbuh. Dengan kata lain, produktivitas jagung dalam model ini dijelaskan oleh kombinasi faktor agroklimat dan tanah, bukan oleh satu variabel saja.



Gambar 5. Feature importance model Random Forest
 (Sumber : Hasil Olahan data menggunakan python)

Hasil explainability ini sangat membantu memperkuat makna praktis model. Model tidak hanya memberi prediksi numerik, tetapi juga mengarahkan perhatian pada faktor yang paling layak diprioritaskan dalam pengelolaan budidaya. Dalam konteks pertanian, informasi semacam ini jauh lebih berguna daripada sekadar mengetahui model mana yang paling akurat, karena hasilnya dapat diterjemahkan ke dalam langkah pengelolaan yang lebih terarah.



6. Pembahasan

Hasil penelitian ini menegaskan bahwa pendekatan machine learning memiliki potensi yang baik untuk memprediksi produktivitas jagung ketika didukung oleh data agroklimat dan karakteristik tanah yang memadai. Kinerja model Random Forest yang paling baik setelah seleksi fitur menunjukkan bahwa hubungan antara produktivitas jagung dan variabel prediktor cenderung nonlinier serta melibatkan interaksi antarfaktor. Hal ini juga menjelaskan mengapa model linier seperti Elastic Net tidak mampu mencapai performa yang setara. Dalam konteks machine learning terapan, temuan ini menguatkan pentingnya pemilihan model yang sesuai dengan struktur data, bukan sekadar memilih model yang paling sederhana.

Dari sisi metodologis, seleksi fitur adaptif terbukti memberi dampak nyata pada performa Random Forest. Penurunan RMSE dari 0,572 menjadi 0,548 menunjukkan bahwa model menjadi lebih presisi setelah fitur yang kurang relevan dieliminasi. Perbaikan ini mungkin tampak moderat secara numerik, tetapi secara substantif penting karena dicapai bersamaan dengan penyederhanaan variabel input. Dengan kata lain, model menjadi lebih efisien sekaligus sedikit lebih akurat. Hasil ini sejalan dengan argumen bahwa seleksi fitur bukan hanya alat reduksi dimensi, tetapi juga strategi untuk memperbaiki fokus model terhadap informasi yang paling bermakna.

Dari sisi substantif, dominasi curah hujan sebagai faktor utama memperlihatkan bahwa air tetap menjadi penentu sentral dalam pembentukan hasil jagung. Akan tetapi, hasil explainability juga menunjukkan bahwa curah hujan tidak berdiri sendiri. pH tanah, suhu, nitrogen, dan penyinaran tetap memberikan kontribusi penting, yang berarti produktivitas jagung lebih tepat dipahami sebagai hasil dari keseimbangan kondisi iklim dan kualitas lahan. Temuan ini selaras dengan perspektif agronomi bahwa respons tanaman terhadap lingkungan bersifat integratif dan sangat bergantung pada kombinasi faktor tumbuh.

Secara praktis, model terbaik dalam penelitian ini dapat dipandang sebagai kerangka awal sistem pendukung keputusan pada budidaya jagung. Informasi mengenai dominasi curah hujan, pH tanah, dan nitrogen dapat digunakan untuk menyusun prioritas intervensi, misalnya pengelolaan irigasi, perbaikan kemasaman tanah, atau penyesuaian dosis pemupukan. Dengan demikian, kontribusi penelitian tidak berhenti pada aspek komputasional, tetapi juga menyentuh sisi praktis dalam pengelolaan produksi jagung berbasis data.

SIMPULAN

Penelitian ini menunjukkan bahwa pendekatan machine learning mampu digunakan secara efektif untuk memprediksi produktivitas jagung berdasarkan data agroklimat dan karakteristik tanah. Hasil analisis menegaskan bahwa hubungan antara produktivitas jagung dan variabel prediktor bersifat kompleks serta tidak sepenuhnya linear, sehingga pemodelan berbasis algoritma nonlinier lebih sesuai dibandingkan pendekatan linier sederhana.

Penerapan seleksi fitur adaptif terbukti mampu menyederhanakan model dengan mempertahankan fitur-fitur yang paling relevan. Dari 10 fitur numerik awal, proses seleksi menghasilkan 6 fitur numerik optimal, yaitu curah hujan, penyinaran, suhu rata-rata, elevasi, nitrogen, dan pH tanah. Hasil ini menunjukkan bahwa tidak semua variabel awal memiliki kontribusi yang sama terhadap model, dan penyederhanaan fitur dapat meningkatkan efisiensi tanpa menghilangkan informasi penting.

Dari sisi performa, Random Forest sesudah seleksi fitur menjadi model terbaik dengan nilai RMSE sebesar 0,548, MAE sebesar 0,426, dan R^2 sebesar 0,528. Hasil ini menunjukkan bahwa seleksi fitur adaptif memberikan dampak positif terhadap kualitas prediksi, khususnya pada model berbasis



pohon. Dengan demikian, kombinasi antara pemilihan fitur yang tepat dan algoritma yang sesuai berperan penting dalam menghasilkan model prediksi yang lebih baik.

Analisis explainable machine learning memperlihatkan bahwa curah hujan merupakan faktor paling dominan dalam memengaruhi produktivitas jagung, diikuti oleh pH tanah, suhu rata-rata, elevasi, penyinaran, dan nitrogen. Temuan ini menegaskan bahwa produktivitas jagung dipengaruhi oleh interaksi antara kondisi iklim dan kualitas tanah. Oleh karena itu, model yang dibangun tidak hanya bermanfaat untuk prediksi, tetapi juga untuk menjelaskan faktor-faktor utama yang perlu diperhatikan dalam pengelolaan budidaya jagung. Penelitian ini menegaskan bahwa integrasi seleksi fitur adaptif dan explainable machine learning merupakan pendekatan yang efektif untuk menghasilkan model prediksi produktivitas jagung yang lebih akurat, efisien, dan interpretatif.

SARAN

Penelitian berikutnya dapat mengembangkan seleksi fitur adaptif yang lebih kuat melalui kombinasi metode filter, wrapper, dan embedded, atau melalui pendekatan optimasi yang lebih lanjut. Selain itu, perbandingan dengan model lain yang lebih interpretabel maupun lebih kompleks perlu dilakukan untuk memperoleh keseimbangan terbaik antara akurasi dan transparansi.

DAFTAR PUSTAKA

- Abdel-salam, M., Kumar, N., & Mahajan, S. (2024). A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning. *Neural Computing and Applications*, 36(33), 20723–20750. <https://doi.org/10.1007/s00521-024-10226-x>
- Agrawal, N., Govil, H., & Kumar, T. (2025). Agricultural land suitability classification and crop suggestion using machine learning and spatial multicriteria decision analysis in semi-arid ecosystem. *Environment, Development and Sustainability*, 27(6), 13689–13726. <https://doi.org/10.1007/s10668-023-04440-1>
- Bhullar, A., Nadeem, K., & Ali, R. A. (2023). Simultaneous multi-crop land suitability prediction from remote sensing data using semi-supervised learning. *Scientific Reports*, 13(1), 6823. <https://doi.org/10.1038/s41598-023-33840-6>
- Croci, M., Impollonia, G., Meroni, M., & Amaducci, S. (2023). Dynamic Maize Yield Predictions Using Machine Learning on Multi-Source Data. *Remote Sensing-MDPI*.
- Debalke, D. B., & Abebe, J. T. (2022). Maize yield forecast using GIS and remote sensing in Kaffa Zone, South West Ethiopia. *Environmental Systems Research*, 11(1), 1. <https://doi.org/10.1186/s40068-022-00249-5>
- Ganati, B. A., & Sitote, T. M. (2025). Predicting land suitability for wheat and barley crops using machine learning techniques. *Scientific Reports*, 15(1), 15879. <https://doi.org/10.1038/s41598-025-99070-0>
- Han, Y., Qiao, D., & Lu, H. (2023). Spatial-temporal coupling pattern between irrigation demand and soil moisture dynamics throughout wheat-maize rotation system in the North China Plain. *European Journal of Agronomy*, 151, 126970. <https://doi.org/https://doi.org/10.1016/j.eja.2023.126970>
- Ismail, I., & Hidayah, A. (2025). Implementasi Machine Learning Dengan Metode Regresi Linear Untuk Prediksi Gaji Karyawan Berdasarkan Masa Kerja. *Jurnal RISTER : Riset Sistem Cerdas*, 2(1), 1–7. <https://doi.org/10.25126/Rister>
- Kang, Y., Ozdogan, M., Zhu, X., Ye, Z., Hain, C., & Anderson, M. (2020). Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US



Midwest OPEN ACCESS Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the . *Environmental Research Letters*.

- Kumar, C., Dhillon, J., Huang, Y., & Reddy, K. (2025). Explainable machine learning models for corn yield prediction using UAV multispectral data. *Computers and Electronics in Agriculture*, *231*, 109990. <https://doi.org/https://doi.org/10.1016/j.compag.2025.109990>
- Mohan, R. N. V. J., Rayanoothala, P. S., & Sree, R. P. (2025). Next-gen agriculture: integrating AI and XAI for precision crop yield predictions. *Frontiers in Plant Science, Volume 15-2024*. <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2024.1451607>
- Nisa, Z. K., Pradipta, A. G., Sholikah, L. N., Pratama, B. F., Prihanantya, A. S., Ngadisih, Susanto, S., & Arif, S. S. (2023). Recognition of Agricultural Land-Use Change with Machine Learning-Based for Regional Food Security Assessment in Kulon Progo Plains Area. *International Journal on Advanced Science, Engineering and Information Technology*, *13*(1), 54–61. <https://doi.org/10.18517/ijaseit.13.1.16550>
- Rajo, H., Santana, D. C., Pereira, L., Teodoro, R., Oliveira, I. C. De, Gava, R., Oliveira, G. De, Antonio, C., Teodoro, P. E., & Shiratsuchi, L. S. (2023). Maize Yield Prediction with Machine Learning , Spectral Variables and Irrigation Management. *Remote Sensing-MDPI*, *15*(1), 79. <https://doi.org/https://doi.org/10.3390/rs15010079>
- Ryo, M. (2022). Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. *Artificial Intelligence in Agriculture*, *6*, 257–265. <https://doi.org/https://doi.org/10.1016/j.aiaa.2022.11.003>
- Santill, A., & Salinas-moreno, Y. (2021). Spatial-Temporal Evolution of Scientific Production about Genetically Modified Maize. *Agriculture (Switzerland)*, 1–14. <https://doi.org/https://doi.org/10.3390/agriculture11030246>
- Sathiyamurthi, S., Subbarayan, S., Ramya, M., Sivasakthi, M., Gobi, R., Qaysi, S., Praveen Kumar, S., Lee, J., Alarifi, N., Wahba, M., & M. Youssef, Y. (2024). Advancing Agricultural Land Suitability in Urbanized Semi-Arid Environments: Insights from Geospatial and Machine Learning Approaches. *ISPRS International Journal of Geo-Information*, *13*(12), 1–23. <https://doi.org/10.3390/ijgi13120436>
- Sharma, R. K., Kaur, J., Feng, G., Huang, Y., Kumar, C., Wang, Y., Sharma, S., Jenkins, J., & Dhillon, J. (2025). Maize and soybean yield prediction using machine learning methods: a systematic literature review. *Discover Agriculture*, *3*(1), 64. <https://doi.org/10.1007/s44279-025-00215-6>
- Shawon, S. M., Ema, F. B., Mahi, A. K., Niha, F. L., & Zubair, H. T. (2025). Crop yield prediction using machine learning: An extensive and systematic literature review. *Smart Agricultural Technology*, *10*, 100718. <https://doi.org/https://doi.org/10.1016/j.atech.2024.100718>
- Sukriadi, S., Adawiah, A., & Ismail, I. (2026). Leveraging machine learning for sustainable smart farming in rural landscapes. *Discover Food*, *6*(1), 168. <https://doi.org/10.1007/s44187-026-00888-y>