

ANALISIS KOMPARATIF ALGORITMA *MACHINE LEARNING* UNTUK PREDIKSI PRODUKSI PADI KABUPATEN BANDUNG

Mohammad Bayu Anggara¹, Wini Fetia Wardhiani^{2*}, Inka Aqila Nurfalqi³

^{1,3} Program Studi Teknik Informatika, Fakultas Teknologi Informasi,
Universitas Bale Bandung, Bandung, Jawa Barat, Indonesia.

^{2*} Program Studi Agribisnis, Fakultas Pertanian,
Universitas Bale Bandung, Bandung, Jawa Barat, Indonesia.

*Corresponding author: winifetia2526@gmail.com

INFO ARTIKEL	ABSTRACT/ABSTRAK
Diterima : 04-05-2026 Direvisi : 25-05-2026 Dipublikasi : 24-07-2026	Comparative Analysis of Machine Learning Algorithms for Rice Production Prediction in Bandung Regency
<i>Keywords: rice production; machine learning; regression; XGBoost; random forest</i>	<i>Rice production is an important indicator for supporting regional food security, including in Bandung Regency. This study aims to compare the performance of four machine learning algorithms, namely Linear Regression, Support Vector Regression (SVR), Random Forest, and XGBoost, in predicting rice production based on harvested area and productivity data from 2018 to 2025. The data were processed through missing value checking, chronological ordering by year, and feature engineering using production lag, rolling mean, and the interaction between harvested area and productivity. Model validation was conducted using Leave-One-Out Cross Validation (LOO-CV) because the dataset was limited to eight annual observations. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). The results show that Linear Regression achieved the best performance, with an MAE of 75,277 tons, RMSE of 207,530 tons, MAPE of 0.24%, and R^2 of 0.9931. Random Forest ranked second with a MAPE of 4.45% and R^2 of 0.5508. XGBoost produced a MAPE of 5.08% and R^2 of 0.4810, while SVR showed the lowest performance with a MAPE of 7.36% and R^2 of -0.1583. The rice production forecast for 2026-2030 indicates a stable trend, ranging from approximately 31.28 to 35.37 million tons depending on the model used. This study provides an initial basis for developing data-driven rice production prediction systems at the regency level. However, these results should be interpreted in the context of data limitations, given that only eight annual observations were used.</i>
Kata Kunci: produksi padi; machine learning; regresi; XGBoost; random forest.	Produksi padi merupakan indikator penting dalam mendukung ketahanan pangan daerah, termasuk di Kabupaten Bandung. Penelitian ini bertujuan membandingkan performa empat algoritma machine learning, yaitu Linear Regression, Support Vector Regression (SVR), Random Forest, dan XGBoost, dalam memprediksi produksi padi berdasarkan data luas panen dan produktivitas periode 2018-2025. Data penelitian diproses melalui pemeriksaan nilai kosong, pengurutan data berdasarkan tahun, serta feature engineering berupa lag produksi, rolling mean, dan interaksi antara luas panen dan produktivitas. Validasi model dilakukan menggunakan <i>Leave-One-Out Cross Validation</i> (LOO-CV) karena jumlah data yang digunakan terbatas, yaitu delapan observasi tahunan. Evaluasi model menggunakan <i>Mean Absolute Error</i> (MAE), <i>Root Mean Squared Error</i> (RMSE), <i>Mean Absolute Percentage Error</i> (MAPE), dan koefisien determinasi (R^2). Hasil penelitian menunjukkan bahwa Linear Regression memperoleh performa terbaik dengan MAE 75.277 ton, RMSE 207.530 ton, MAPE 0,24%, dan R^2 0,9931. Random Forest berada pada urutan kedua dengan MAPE 4,45% dan R^2 0,5508. XGBoost menghasilkan MAPE 5,08% dan R^2 0,4810, sedangkan SVR menunjukkan performa terendah dengan MAPE 7,36% dan R^2 -0,1583. Prediksi produksi padi periode 2026-2030 menunjukkan kecenderungan stabil pada kisaran 31,28-35,37 juta ton, bergantung pada model yang digunakan. Hasil penelitian ini dapat menjadi dasar awal dalam pengembangan sistem prediksi produksi padi berbasis data di tingkat kabupaten. Namun, hasil ini perlu dipahami dalam konteks keterbatasan data, mengingat jumlah observasi yang digunakan hanya delapan data tahunan.

PENDAHULUAN

Sektor pertanian memiliki peranan penting dalam pembangunan ekonomi Indonesia, khususnya dalam penyediaan pangan nasional dan penyerapan tenaga kerja. Salah satu komoditas strategis yang menjadi perhatian utama pemerintah adalah padi,

karena beras merupakan makanan pokok sebagian besar masyarakat Indonesia. Stabilitas produksi padi sangat menentukan ketahanan pangan nasional, sehingga diperlukan upaya yang mampu meningkatkan efektivitas pengelolaan produksi dan

pengambilan keputusan di sektor pertanian. Kabupaten Bandung sebagai salah satu daerah penghasil padi di Provinsi Jawa Barat memiliki kontribusi penting terhadap produksi pangan regional. Namun, produktivitas padi di wilayah ini masih menghadapi berbagai tantangan, seperti perubahan iklim, alih fungsi lahan, serangan hama dan penyakit, serta ketidakpastian kondisi cuaca yang dapat memengaruhi hasil produksi setiap musim tanam.

Perkembangan teknologi informasi dan transformasi digital pada sektor pertanian telah membuka peluang baru dalam pengolahan dan analisis data pertanian. Salah satu teknologi yang berkembang pesat adalah *machine learning*, yaitu cabang dari kecerdasan buatan (*Artificial Intelligence*) yang mampu mempelajari pola dari data historis untuk menghasilkan prediksi yang lebih akurat. Dalam bidang pertanian, penerapan *machine learning* mulai banyak digunakan untuk mendukung sistem pengambilan keputusan, termasuk prediksi hasil panen, identifikasi penyakit tanaman, analisis cuaca, hingga pengelolaan rantai pasokan hasil pertanian. Kemampuan algoritma *machine learning* dalam mengolah data dalam jumlah besar dinilai lebih efektif dibandingkan metode statistik konvensional, terutama pada data yang memiliki pola kompleks dan dinamis.

Prediksi produksi padi menjadi salah satu aspek penting dalam perencanaan pertanian karena dapat membantu pemerintah, petani, maupun pemangku kepentingan lainnya dalam menentukan kebijakan pangan, distribusi hasil panen, serta strategi pengelolaan sumber daya pertanian. Selama ini, prediksi produksi padi umumnya dilakukan menggunakan pendekatan statistik sederhana yang memiliki keterbatasan dalam menangkap hubungan nonlinier antarvariabel.

Padi merupakan salah satu komoditas strategis karena berhubungan langsung dengan ketersediaan pangan masyarakat. Pada tingkat daerah, informasi mengenai produksi padi dibutuhkan untuk memperkirakan kebutuhan distribusi, menjaga stabilitas pasokan, dan membantu penyusunan kebijakan pertanian. Masalahnya, produksi padi tidak selalu bergerak secara tetap dari tahun ke tahun. Perubahan luas panen, produktivitas lahan, kondisi lingkungan, serta kejadian tertentu dapat menyebabkan produksi meningkat atau menurun pada periode tertentu.

Perkembangan analisis data membuat prediksi produksi pertanian tidak hanya dilakukan dengan pengamatan manual, tetapi juga melalui model berbasis *machine learning*. Pendekatan ini dapat mempelajari hubungan antara data historis dan variabel penjelas, kemudian menghasilkan estimasi untuk periode berikutnya. van Klompenburg et al., (2020) menjelaskan bahwa *machine learning* banyak digunakan dalam prediksi hasil tanaman karena dapat mendukung perencanaan pertanian dan

pengambilan keputusan berbasis data. Shawon et al., (2025) juga menunjukkan bahwa penggunaan *machine learning* pada prediksi hasil tanaman terus berkembang karena mampu mengolah pola data pertanian yang beragam. Javed & Azmi Murad, (2024) menekankan bahwa prediksi hasil pertanian membutuhkan data yang relevan, seperti luas panen, kondisi cuaca, dan indikator produktivitas. Paudel et al., (2021) menambahkan bahwa data pertanian berskala besar dapat dimanfaatkan untuk menghasilkan prediksi hasil panen yang lebih mendukung proses pengambilan keputusan.

Beberapa penelitian terdahulu menunjukkan bahwa setiap algoritma dapat menghasilkan performa yang berbeda karena cara kerja modelnya tidak sama. Choudhary et al., (2022) menggunakan Random Forest untuk pemetaan dan prediksi hasil padi berbasis data Sentinel-2, sehingga metode ini relevan untuk data pertanian yang memiliki hubungan non-linear. Manurung et al., (2025) juga menerapkan Random Forest Regressor pada prediksi produksi padi di Indonesia dan menunjukkan bahwa model tersebut dapat digunakan untuk membaca hubungan antara variabel pertanian dan produksi. Chen & Guestin (2016) menjelaskan bahwa XGBoost bekerja dengan teknik boosting, yaitu memperbaiki kesalahan prediksi secara bertahap dari model sebelumnya. Crane-Droesch, (2018) menunjukkan bahwa pendekatan *machine learning* dapat digunakan dalam prediksi hasil tanaman dan penilaian dampak iklim pada sektor pertanian. Yasin et al., (2025) memperlihatkan bahwa model linear tetap dapat memberikan hasil baik pada prediksi produksi padi apabila hubungan antara variabel input dan target relatif sederhana.

Penelitian terkait prediksi produksi padi di Indonesia telah dilakukan dengan berbagai pendekatan. Putra & Ulfa Walmi, (2020) menerapkan Artificial Neural Network untuk prediksi produksi padi dan menilai hasilnya dengan MAPE. Cahyono & Rosadi, (2025) menunjukkan bahwa ANN dapat menangkap hubungan non-linear pada data produksi padi berbasis faktor agroklim. Erlin et al., (2024) juga membandingkan beberapa model *machine learning* untuk prediksi produksi padi Indonesia dengan memanfaatkan data pertanian dan iklim. Namun, kajian yang berfokus pada data tingkat Kabupaten Bandung masih terbatas, terutama dengan membandingkan beberapa algoritma pada dataset tahunan berukuran kecil. Penggunaan *machine learning* pada dataset kecil tetap dapat dilakukan dengan pendekatan validasi yang tepat, seperti *Leave-One-Out Cross Validation* (LOO-CV), rekayasa fitur temporal, serta pemilihan model yang disesuaikan dengan karakteristik data (Yasin et al., 2025).

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk: (1) membangun model prediksi produksi padi Kabupaten Bandung menggunakan Linear Regression, SVR, Random Forest, dan

XGBoost; (2) membandingkan performa model menggunakan MAE, RMSE, MAPE, dan R^2 ; serta (3) menghasilkan estimasi produksi padi tahun 2026-2030 sebagai gambaran awal bagi perencanaan pertanian daerah. Perbandingan dilakukan berdasarkan tingkat akurasi, nilai kesalahan prediksi, dan kemampuan model dalam mengolah data produksi padi. Hasil penelitian diharapkan dapat memberikan kontribusi dalam pengembangan sistem prediksi produksi pertanian yang lebih akurat serta menjadi bahan pertimbangan bagi pemerintah daerah dan pelaku sektor pertanian dalam mendukung perencanaan dan pengambilan keputusan di bidang pangan.

BAHAN DAN METODE

Tabel 1. Statistik deskriptif dataset produksi padi Kabupaten Bandung tahun 2018-2025

Variabel	Min	Maks	Rata-rata	Std. Deviasi
Produksi padi (ton)	27.715.628	35.330.700	31.547.490	2.678.994
Luas panen (ha)	4.753.779	5.631.025	5.201.770	319.935
Produktivitas (ku/ha)	5.527	6.433	6.062	288

Alur Penelitian

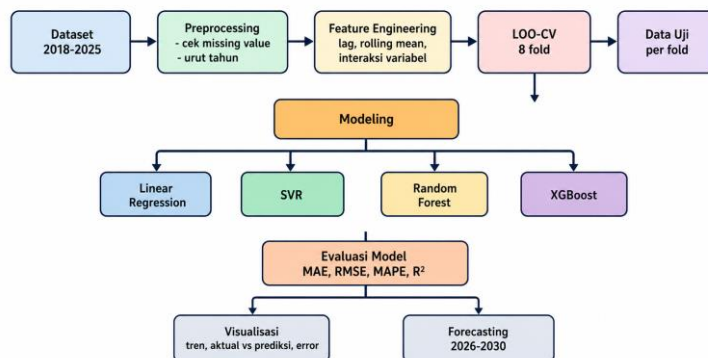
Alur penelitian disusun agar proses pemodelan dapat diulang secara sistematis. Tahapan dimulai dari input dataset produksi padi tahun 2018-2025, kemudian dilanjutkan dengan pemeriksaan data, prapemrosesan, rekayasa fitur, validasi LOO-CV,

Data Penelitian

Penelitian ini menggunakan data sekunder produksi padi Kabupaten Bandung periode 2018–2025 yang diperoleh dari publikasi tabel statistik Badan Pusat Statistik (BPS) Provinsi Jawa Barat berjudul Luas Panen, Produktivitas, dan Produksi Padi Menurut Kabupaten/Kota di Provinsi Jawa Barat. Data terdiri atas empat kolom, yaitu tahun, produksi_padi, luas_panen, dan produktivitas_ku_ha. Variabel produksi_padi digunakan sebagai variabel target, sedangkan luas_panen dan produktivitas_ku_ha digunakan sebagai variabel prediktor. Jumlah observasi sebanyak delapan data tahunan, sehingga metode validasi dipilih dengan mempertimbangkan ukuran dataset yang kecil.

pemodelan menggunakan empat algoritma, evaluasi model, forecasting tahun 2026-2030, dan penyajian output berupa grafik serta tabel hasil. Alur penelitian ditunjukkan pada Gambar 1.

Alur Penelitian Prediksi Produksi Padi Kabupaten Bandung



Gambar 1. Alur penelitian prediksi produksi padi

Prapemrosesan Data

Prapemrosesan dilakukan dengan memastikan kolom tahun tersusun secara kronologis, memeriksa tipe data numerik, menghapus duplikasi tahun apabila ada, serta mengecek nilai kosong. Pemeriksaan nilai kosong penting dilakukan karena data yang tidak lengkap dapat memengaruhi hasil pelatihan model. Pada dataset ini, seluruh variabel utama dapat digunakan dalam proses pemodelan. Untuk model yang sensitif terhadap skala data, fitur numerik disiapkan dengan penskalaan agar nilai antarvariabel tidak saling mendominasi.

Rekayasa Fitur

Rekayasa fitur dilakukan untuk menambah informasi temporal pada dataset. Fitur yang digunakan meliputi lag produksi satu tahun sebelumnya, lag produksi dua tahun sebelumnya, rolling mean dua tahun, indeks tahun, serta interaksi antara luas panen dan produktivitas. Fitur lag dan rolling mean digunakan agar model memperoleh informasi produksi pada periode sebelumnya, sedangkan fitur interaksi digunakan karena produksi padi secara praktis berhubungan dengan luas panen dan tingkat produktivitas. Pada baris awal yang belum memiliki nilai lag, dilakukan penanganan

nilai kosong agar seluruh observasi tetap dapat digunakan.

Algoritma yang Digunakan

Penelitian ini membandingkan empat algoritma. Linear Regression digunakan sebagai model dasar untuk melihat hubungan linear antarvariabel. SVR digunakan untuk menangani pola non-linear dengan pendekatan kernel, tetapi hasilnya dapat dipengaruhi oleh pemilihan parameter. Random Forest digunakan karena mampu menggabungkan banyak pohon keputusan sehingga prediksi menjadi lebih stabil. XGBoost digunakan karena proses boosting-nya dapat memperbaiki kesalahan model secara bertahap dan sering memberikan performa baik pada data tabular. Chen & Guestrin, (2016) menjelaskan bahwa XGBoost dirancang sebagai sistem tree boosting yang efisien dan skalabel. Erlin et al., (2024) membandingkan beberapa model machine learning untuk prediksi produksi padi Indonesia dengan memanfaatkan data pertanian dan iklim. Yasin et al., (2025) juga membandingkan model Linear

Regression, SVR, Random Forest, dan XGBoost dalam konteks prediksi produksi padi multivariat.

Validasi dan Evaluasi Model

Validasi model menggunakan *Leave-One-Out Cross Validation* (LOO-CV). Metode ini dipilih karena dataset hanya berisi delapan observasi, sehingga setiap tahun digunakan bergantian sebagai data uji dan sisa tahun lainnya sebagai data latih. Dengan cara ini, seluruh observasi memperoleh kesempatan untuk diuji. Evaluasi dilakukan menggunakan MAE, RMSE, MAPE, dan R^2 . MAE menunjukkan rata-rata selisih absolut, RMSE memberi penalti lebih besar pada error yang tinggi, MAPE menunjukkan persentase error, sedangkan R^2 menunjukkan kemampuan model menjelaskan variasi data aktual. Pada interpretasi MAPE, nilai di bawah 10% dapat dikategorikan sangat baik, 10-20% baik, 20-50% cukup, dan di atas 50% buruk sebagaimana digunakan dalam penelitian prediksi produksi padi berbasis ANN (Putra & Ulfa Walmi, 2020).

Tabel 2 Metrik evaluasi model

Metrik	Rumus	Keterangan
MAE	$\sum y - \hat{y} / n$	Rata-rata selisih absolut aktual dan prediksi
RMSE	$\sqrt{\sum (y - \hat{y})^2 / n}$	Akar rata-rata kuadrat error
MAPE	$\sum (y - \hat{y} / y) / n \times 100\%$	Persentase rata-rata error absolut
R^2	$1 - SSE/SST$	Proporsi variasi target yang dijelaskan model

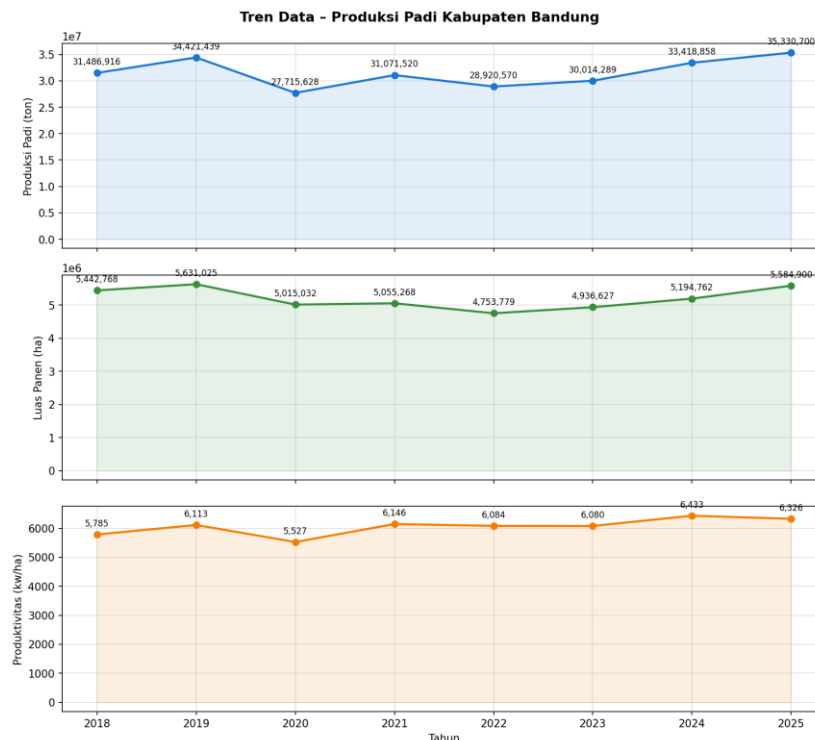
HASIL DAN PEMBAHASAN

Analisis Tren Data

Data yang digunakan dalam penelitian ini terdiri atas produksi padi, luas panen, dan produktivitas Kabupaten Bandung pada periode 2018-2025. Berdasarkan Gambar 2, produksi padi menunjukkan pola naik-turun selama periode pengamatan. Produksi tertinggi terjadi pada tahun 2025 sebesar 35.330.700 ton, sedangkan produksi terendah terjadi pada tahun 2020 sebesar 27.715.628 ton. Setelah penurunan pada tahun 2020, produksi kembali

meningkat secara bertahap hingga akhir periode pengamatan.

Pola luas panen juga bergerak fluktuatif. Luas panen tertinggi berada pada tahun 2019 sebesar 5.631.025 ha, kemudian turun hingga mencapai nilai terendah pada tahun 2022 sebesar 4.753.779 ha. Pada tahun 2023-2025, luas panen kembali meningkat. Produktivitas berada pada rentang 5.527-6.433 ku/ha, dengan nilai terendah pada tahun 2020 dan nilai tertinggi pada tahun 2024. Perubahan ketiga variabel tersebut menunjukkan bahwa model prediksi perlu membaca hubungan antara produksi, luas panen, dan produktivitas secara bersamaan, bukan hanya melihat tren produksi saja.



Gambar 2. Tren data produksi padi, luas panen, dan produktivitas Kabupaten Bandung tahun 2018-2025

Perbandingan Performa Model

Evaluasi model dilakukan menggunakan *Leave-One-Out Cross Validation (LOO-CV)*. Hasil evaluasi disajikan pada Tabel 3. Berdasarkan hasil tersebut, Linear Regression memperoleh performa terbaik dengan MAE 75.277 ton, RMSE 207.530 ton, MAPE 0,24%, dan R^2 0,9931. Nilai MAPE tersebut termasuk kategori sangat baik. Hasil ini menunjukkan bahwa pada dataset berukuran kecil dan memiliki hubungan fitur yang kuat, model sederhana dapat memberikan prediksi yang stabil. Random Forest berada pada posisi kedua dengan MAE 1.392.931 ton, RMSE 1.679.512 ton, MAPE 4,45%, dan R^2 0,5508. XGBoost berada pada posisi ketiga dengan MAE 1.595.736 ton, RMSE 1.805.352 ton, MAPE 5,08%, dan R^2 0,4810. Kedua model tersebut masih masuk kategori sangat baik berdasarkan MAPE, tetapi nilai R^2 -nya lebih rendah dibandingkan Linear Regression. Kondisi ini menunjukkan bahwa model ensemble belum memperoleh variasi data yang cukup untuk membentuk pola percabangan dan boosting secara

optimal pada dataset yang hanya terdiri atas delapan observasi.

Choudhary et al., (2022) menunjukkan bahwa Random Forest dapat bekerja baik untuk prediksi hasil padi ketika data yang digunakan memiliki cakupan spasial dan fitur yang lebih kaya. Manurung et al., (2025) juga melaporkan bahwa Random Forest Regressor efektif digunakan pada prediksi produksi padi Indonesia dengan kombinasi variabel pertanian dan iklim. Erlin et al., (2024) memperlihatkan bahwa model machine learning lanjutan lebih optimal ketika didukung oleh jumlah data dan ragam variabel yang lebih besar. SVR memperoleh performa terendah dengan MAE 2.324.656 ton, RMSE 2.697.075 ton, MAPE 7,36%, dan R^2 -0,1583. Walaupun nilai MAPE masih berada dalam kategori sangat baik, nilai R^2 negatif menunjukkan bahwa SVR belum mampu menjelaskan variasi data aktual dengan baik. Pada grafik prediksi, SVR cenderung menghasilkan nilai yang hampir konstan di sekitar 31,28 juta ton, sehingga kurang mampu mengikuti titik produksi yang tinggi maupun rendah.

Tabel 3 Hasil evaluasi model menggunakan LOO-CV

Model	MAE (ton)	RMSE (ton)	MAPE (%)	R^2
Linear Regression	75.277	207.530	0,24	0,9931
Random Forest	1.392.931	1.679.512	4,45	0,5508
XGBoost	1.595.736	1.805.352	5,08	0,4810
SVR	2.324.656	2.697.075	7,36	-0,1583

Perbandingan Aktual dan Prediksi In-Sample

Perbandingan aktual dan prediksi pada Tabel 4 bersifat in-sample. Artinya, grafik dan tabel ini

digunakan untuk melihat kedekatan pola antara data aktual dan hasil model setelah model mempelajari data historis. Oleh karena itu, dasar utama penentuan

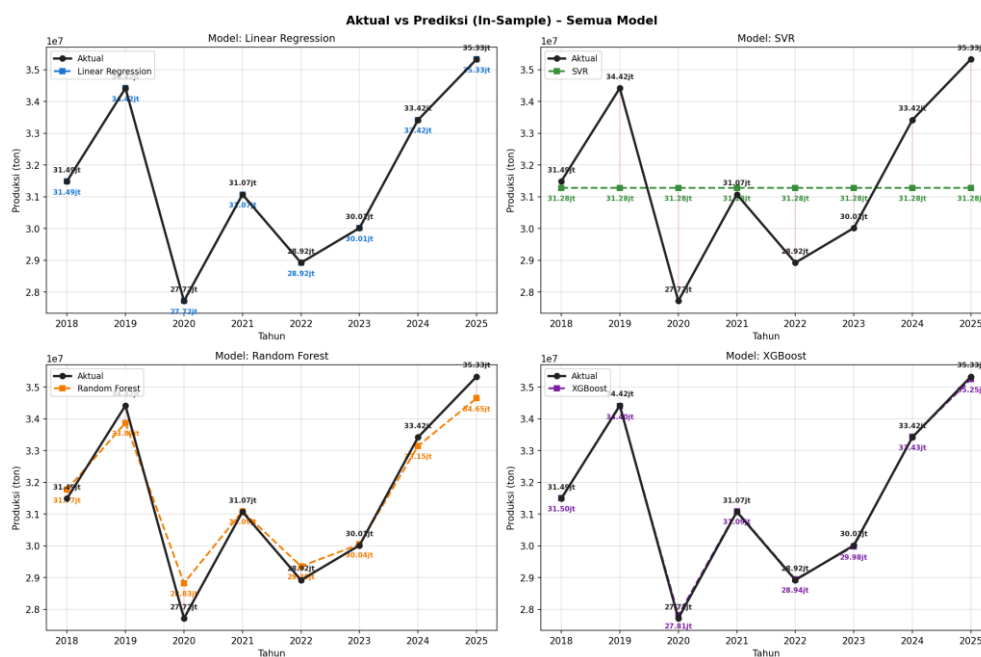
performa model tetap menggunakan hasil LOO-CV pada Tabel 4.

Tabel 4 Perbandingan nilai aktual dan prediksi in-sample

Tahun	Aktual	Linear Regression	SVR	Random Forest	XGBoost
2018	31.486.916	31.486.916	31.279.242	31.771.894	31.499.372
2019	34.421.439	34.421.439	31.279.270	33.874.486	34.401.340
2020	27.715.628	27.715.628	31.279.193	28.825.147	27.809.154
2021	31.071.520	31.071.520	31.279.194	31.091.241	31.091.610
2022	28.920.570	28.920.570	31.279.190	29.354.820	28.944.574
2023	30.014.289	30.014.289	31.279.189	30.044.891	29.983.186
2024	33.418.858	33.418.858	31.279.245	33.146.042	33.434.582
2025	35.330.700	35.330.700	31.279.289	34.648.886	35.250.548

Perbandingan aktual dan prediksi pada Tabel 4 bersifat in-sample, sehingga digunakan untuk melihat kedekatan pola antara data aktual dan hasil model, bukan sebagai satu-satunya dasar penentuan akurasi. Linear Regression mengikuti nilai aktual paling dekat pada hampir seluruh tahun pengamatan. Random Forest dan XGBoost juga mampu

mengikuti pola naik-turun produksi, tetapi masih memiliki selisih pada tahun tertentu, terutama ketika produksi berubah cukup tajam. SVR terlihat lebih datar sehingga kurang mampu menangkap titik produksi tinggi dan rendah. Visualisasi pada Gambar 5 membantu memperlihatkan perbedaan pola prediksi keempat algoritma secara lebih jelas.



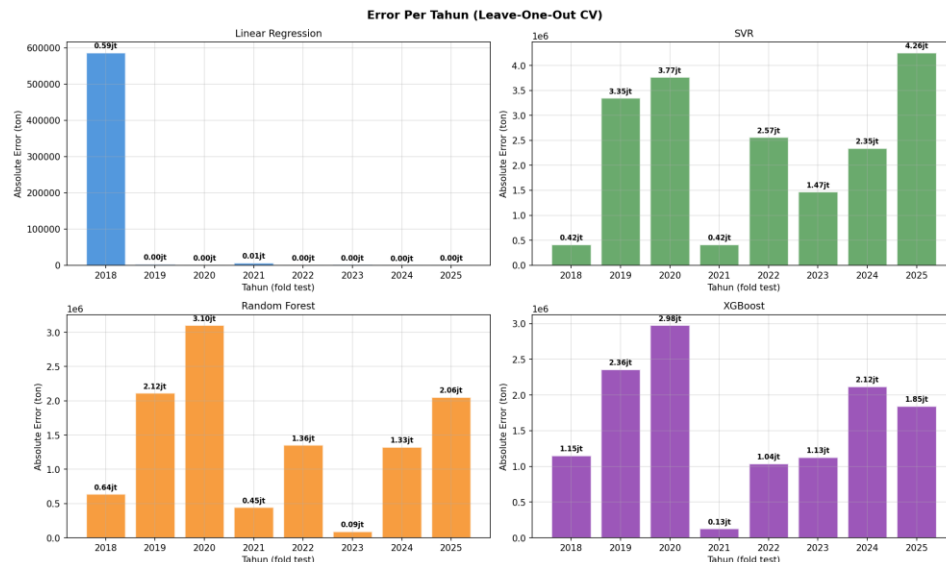
Gambar 3. Grafik aktual dan prediksi in-sample untuk empat algoritma

Error Per Tahun pada LOO-CV

Gambar 4 menunjukkan error absolut pada setiap tahun pengujian LOO-CV. Linear Regression memiliki error paling kecil pada hampir seluruh tahun, meskipun terdapat error yang lebih terlihat pada tahun 2018. Random Forest dan XGBoost menghasilkan error yang lebih besar pada tahun-tahun dengan perubahan produksi cukup tajam, terutama tahun 2020. SVR memiliki error relatif besar pada beberapa tahun karena hasil prediksinya

cenderung mendekati nilai rata-rata dan tidak mengikuti fluktuasi produksi aktual.

Temuan ini memperkuat hasil pada Tabel 3. Model yang mampu mengikuti perubahan data aktual secara lebih baik akan menghasilkan MAE, RMSE, dan MAPE yang lebih rendah. Sebaliknya, model yang menghasilkan prediksi terlalu datar akan memiliki R^2 rendah atau bahkan negatif karena tidak mampu menjelaskan variasi data aktual.



Gambar 4. Error absolut per tahun berdasarkan Leave-One-Out Cross Validation

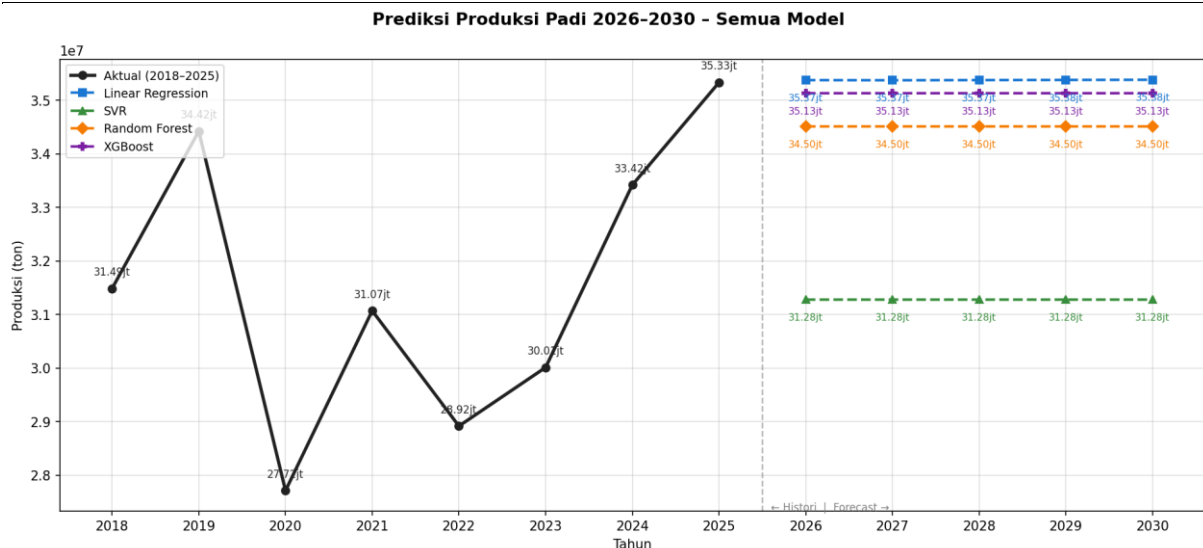
Prediksi Produksi Padi Tahun 2026-2030

Prediksi produksi padi tahun 2026-2030 disajikan pada Tabel 5 dan Gambar 5. Linear Regression memperkirakan produksi padi Kabupaten Bandung berada pada kisaran 35,37-

35,38 juta ton. XGBoost memperkirakan produksi sekitar 35,13 juta ton, sedangkan Random Forest memperkirakan produksi sekitar 34,50 juta ton. SVR memberikan prediksi paling rendah dan relatif konstan, yaitu sekitar 31,28 juta ton.

Tabel 5 Prediksi produksi padi Kabupaten Bandung tahun 2026-2030

Tahun	Linear Regression	XGBoost	Random Forest	SVR
2026	35.374.222	35.132.400	34.504.769	31.279.284
2027	35.371.620	35.132.400	34.504.769	31.279.262
2028	35.374.571	35.132.400	34.504.769	31.279.261
2029	35.377.543	35.132.400	34.504.769	31.279.261
2030	35.380.515	35.132.400	34.504.769	31.279.260



Gambar 5. Prediksi produksi padi Kabupaten Bandung tahun 2026-2030

Berdasarkan Tabel 5, Linear Regression memperkirakan produksi padi Kabupaten Bandung pada periode 2026-2030 berada di sekitar 35,37-35,38 juta ton. XGBoost memperkirakan produksi sekitar 35,13 juta ton, Random Forest sekitar 34,50 juta ton, dan SVR sekitar 31,28 juta ton. Perbedaan

hasil antar model menunjukkan bahwa pemilihan algoritma berpengaruh terhadap angka prediksi akhir. Karena Linear Regression menjadi model terbaik berdasarkan hasil evaluasi LOO-CV, hasil dari model tersebut dapat dijadikan acuan utama,

Pembahasan Umum

Hasil penelitian ini memperlihatkan bahwa model dengan performa terbaik tidak selalu model yang paling kompleks. Pada dataset kecil, Linear Regression dapat bekerja baik karena hubungan antara produksi padi, luas panen, produktivitas, dan fitur hasil rekayasa masih dapat dijelaskan secara sederhana. Islam et al., (2024) menunjukkan bahwa perbandingan beberapa model machine learning dapat membantu menentukan model yang paling sesuai dengan karakteristik data hasil pertanian tertentu.

Liu et al., (2022) menggunakan indikator satelit dan iklim untuk prediksi hasil padi, sehingga penambahan variabel eksternal dapat memperkaya informasi model. Oikonomidis et al., (2023) menjelaskan bahwa model deep learning untuk prediksi hasil tanaman umumnya membutuhkan data yang lebih besar agar proses pembelajaran pola dapat berjalan optimal. Dengan demikian, hasil penelitian ini perlu dipahami dalam konteks dataset yang terbatas, yaitu delapan data tahunan.

Terkait horizon prediksi 2026–2030 yang digunakan dalam penelitian ini, pemilihan rentang lima tahun ke depan didasarkan pada relevansinya dengan siklus perencanaan pangan jangka menengah di tingkat kabupaten. Prediksi dilakukan secara iteratif dengan memanfaatkan nilai prediksi sebelumnya sebagai fitur lag pada langkah berikutnya. Mengingat dataset hanya terdiri atas delapan observasi tahunan, hasil prediksi 2026–2030 sebaiknya dipahami sebagai estimasi awal yang bersifat indikatif, bukan proyeksi final. Validasi empiris menggunakan data aktual masa depan tetap diperlukan untuk mengonfirmasi ketepatan prediksi ini (Oikonomidis et al., 2023).

Kemungkinan overfitting akibat feature engineering telah dipertimbangkan dalam penelitian ini. Meskipun fitur lag dan rolling mean diturunkan dari variabel target sehingga berpotensi menghasilkan korelasi yang tinggi secara in-sample, evaluasi utama menggunakan LOO-CV yang secara ketat memisahkan data latih dan uji pada setiap iterasi. Hasil evaluasi LOO-CV menunjukkan nilai MAPE yang masih berada dalam kategori sangat baik untuk seluruh model (0,24%–7,36%), dan nilai R^2 yang positif untuk tiga dari empat model, sehingga indikasi overfitting yang signifikan tidak ditemukan dalam penelitian ini.

Keterbatasan utama penelitian ini terletak pada jumlah data yang masih sedikit. Jumlah tersebut cukup untuk membuat prototipe prediksi, tetapi belum cukup kuat untuk menyimpulkan pola jangka panjang. Oleh karena itu, hasil prediksi 2026–2030 sebaiknya dipahami sebagai estimasi awal, bukan angka final. Penelitian berikutnya disarankan menambah rentang tahun, memasukkan variabel iklim seperti curah hujan, suhu, dan kelembapan, serta melakukan tuning parameter yang lebih

mendalam pada SVR, Random Forest, dan XGBoost.

KESIMPULAN

Penelitian ini membandingkan Linear Regression, SVR, Random Forest, dan XGBoost untuk memprediksi produksi padi Kabupaten Bandung menggunakan data tahunan 2018–2025. Hasil evaluasi LOO-CV menunjukkan bahwa Linear Regression menjadi model terbaik dengan MAE 75.277 ton, RMSE 207.530 ton, MAPE 0,24%, dan R^2 0,9931. Random Forest berada pada posisi kedua dengan MAPE 4,45% dan R^2 0,5508, sedangkan XGBoost berada pada posisi ketiga dengan MAPE 5,08% dan R^2 0,4810. SVR memiliki performa terendah dengan MAPE 7,36% dan R^2 negatif. Prediksi tahun 2026–2030 berdasarkan model terbaik menunjukkan produksi padi relatif stabil pada kisaran 31,28–35,37 juta ton. Temuan ini menunjukkan bahwa model sederhana dapat menjadi pilihan yang tepat ketika dataset kecil dan hubungan antarvariabel relatif kuat. Penelitian lanjutan perlu menggunakan data historis yang lebih panjang serta menambahkan faktor iklim agar hasil prediksi lebih kuat untuk mendukung perencanaan pangan daerah.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan bantuan dalam pelaksanaan penelitian ini. Ucapan terima kasih disampaikan kepada Pemerintah Kabupaten Bandung dan instansi terkait yang telah menyediakan data pendukung penelitian. Penulis juga menyampaikan apresiasi kepada pihak akademik dan rekan-rekan dosen yang telah memberikan masukan, arahan, serta dukungan selama proses penelitian dan penyusunan artikel ini. Selain itu, penulis mengucapkan terima kasih kepada keluarga dan seluruh pihak yang tidak dapat disebutkan satu per satu atas doa, motivasi, dan bantuan yang diberikan sehingga penelitian mengenai analisis komparatif algoritma machine learning untuk prediksi produksi padi di Kabupaten Bandung ini dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-August-2016, 785–794.
<https://doi.org/10.1145/2939672.2939785>
- Choudhary, K., Shi, W., Dong, Y., & Paringer, R. (2022). Random Forest for rice yield mapping and prediction using Sentinel-2 data with

- Google Earth Engine. *Advances in Space Research*, 70(8), 2443–2457.
<https://doi.org/10.1016/J.ASR.2022.06.073>
- Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, 13(11).
<https://doi.org/10.1088/1748-9326/aae159>
- Erlin, Yuniarta, A., Wulandhari, L. A., Desnelita, Y., Nasution, N., & Junadhi. (2024). Enhancing Rice Production Prediction in Indonesia Using Advanced Machine Learning Models. *IEEE Access*, 12, 151161–151177.
<https://doi.org/10.1109/ACCESS.2024.3478738>
- Islam, T., Mazumder, T., Shahriair Roni, M. N., & Nur, M. S. (2024). A comparative study of machine learning models for predicting Aman rice yields in Bangladesh. *Heliyon*, 10(23).
<https://doi.org/10.1016/j.heliyon.2024.e40764>
- Jabed, M. A., & Azmi Murad, M. A. (2024). Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. In *Heliyon* (Vol. 10, Issue 24). Elsevier Ltd.
<https://doi.org/10.1016/j.heliyon.2024.e40836>
- Liu, Y., Wang, S., Chen, J., Chen, B., Wang, X., Hao, D., & Sun, L. (2022). Rice Yield Prediction and Model Interpretation Based on Satellite and Climatic Indicators Using a Transformer Method. *Remote Sensing*, 14(19). <https://doi.org/10.3390/rs14195045>
- Manurung, D., Zealtiel, B., & Lubis, A. H. (2025). Prediksi Produksi Tanaman Padi di Indonesia dengan Menggunakan Algoritma Random Forest Regressor. *Journal of Computing and Informatics Research*, 4(3), 345.
<https://doi.org/10.47065/comforch.v4i3.2125>
- Oikonomidis, A., Catal, C., & Kassahun, A. (2023). Deep learning for crop yield prediction: a systematic literature review. In *New Zealand Journal of Crop and Horticultural Science* (Vol. 51, Issue 1, pp. 1–26). Taylor and Francis Ltd.
- Pusat Statistik Provinsi Jawa Barat. <https://doi.org/10.1080/01140671.2022.2032213>
- Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., & Athanasiadis, I. N. (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, 187, 103016.
<https://doi.org/10.1016/J.AGSY.2020.103016>
- Putra, H., & Ulfa Walmi, N. (2020). Penerapan Prediksi Produksi Padi Menggunakan Artificial Neural Network Algoritma Backpropagation. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 6(2).
<https://doi.org/10.25077/teknosi.v6i2.2020.100-107>
- Setiawan Cahyono, & Muhammad Imron Rosadi. (2025). Penerapan Artificial Neural Network untuk Prediksi Produksi Padi di Sumatera. *Jurnal Informatika Polinema*, 11(4).
<https://doi.org/10.33795/jip.v11i4.7727>
- 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/10.1016/J.ATECH.2024.100718>
- van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177, 105709.
<https://doi.org/10.1016/J.COMPAG.2020.105709>
- Yasin, F., Firmansyah, M. R., Aldo, D., & Amrustian, M. A. (2025). Multivariate Forecasting of Paddy Production: A Comparative Study of Machine Learning Models. *Jurnal Teknik Informatika (Jutif)*, 6(3), 1431–1442.
<https://doi.org/10.52436/1.jutif.2025.6.3.4681>
- Badan Pusat Statistik Provinsi Jawa Barat. (2025). *Luas panen, produktivitas, dan produksi padi menurut kabupaten/kota di Provinsi Jawa Barat, 2025*. Bad