

PREDIKSI PRODUKSI PADI KABUPATEN KARAWANG BERDASARKAN LUAS PANEN PADI MENGGUNAKAN CITRA SATELIT

¹Rizky Afriza Rosal Fino, ²Aries Suharso, ³Chaerur Rozikin
¹⁻³Universitas Singaperbangsa Karawang, Kabupaten Karawang

E-mail: ¹2210631170148@student.unsika.ac.id, ²aries.suharso@staff.unsika.ac.id,
³chaerur.rozikin@staff.unsika.ac.id

ABSTRAK

Kabupaten Karawang merupakan salah satu lumbung padi utama di Provinsi Jawa Barat, tetapi alih fungsi lahan dan dinamika iklim membuat pemantauan produksi padi yang akurat dan tepat waktu menjadi kebutuhan yang semakin mendesak. Penelitian ini bertujuan membangun dan mengevaluasi model prediksi produktivitas padi (ton/ha) berbasis citra satelit multisensor dan machine learning pada 30 kecamatan di Kabupaten Karawang selama periode 2019 sampai 2025, dengan mengadopsi kerangka CRISP-DM. Fitur masukan diekstraksi dari citra Sentinel-2 (NDVI dan LSWI), Sentinel-1 SAR, serta data iklim pendukung melalui Google Earth Engine. Lima algoritma diuji secara komparatif menggunakan skema validasi Leave-One-Year-Out (LOYO), yaitu Random Forest, Ordinary Least Squares (OLS), Ridge Regression, XGBoost, dan LightGBM. Random Forest menjadi model terbaik pada level kecamatan dengan MAPE 7,207%, RMSE 0,580 ton/ha, dan Akurasi Kabupaten 94,154%, dengan R^2 sebesar 0,961 pada skala produksi absolut tingkat kecamatan. Pada level kabupaten, OLS tampil lebih unggul dengan MAPE 5,279%. Saat diterapkan untuk estimasi produksi tahun 2025, model menghasilkan prediksi 1.475.946,3 ton dibandingkan data aktual 1.439.910,2 ton, setara Akurasi Kabupaten 97,50%. Temuan ini menegaskan bahwa model berbasis citra satelit dapat difungsikan sebagai instrumen pendukung estimasi produktivitas dan produksi padi di Kabupaten Karawang..

Kata kunci : Prediksi Produktivitas Padi, Citra Satelit Sentinel, NDVI Deret Waktu, Machine Learning, Leave-One-Year-Out, Kabupaten Karawang.

ABSTRACT

Karawang Regency is one of West Java's primary rice-producing regions, yet land-use conversion and climate variability have made accurate, timely monitoring of rice production increasingly critical. This study develops and evaluates a rice productivity prediction model (ton/ha) based on multi-sensor satellite imagery and machine learning across 30 sub-districts in Karawang Regency over 2019–2025, adopting the CRISP-DM framework. Features were extracted from Sentinel-2 imagery (NDVI and LSWI), Sentinel-1 SAR, and climate data via Google Earth Engine. Five algorithms were compared using Leave-One-Year-Out (LOYO) validation: Random Forest, Ordinary Least Squares (OLS), Ridge Regression, XGBoost, and LightGBM. Random Forest was the best-performing model at the sub-district level, with a MAPE of 7.207%, RMSE of 0.580 ton/ha, and Regency Accuracy of 94.154%, with an R^2 of 0.961 on the absolute production scale. At the regency level, OLS performed better, with a MAPE of 5.279%. Applied to estimate 2025 production, the model produced 1,475,946.3 tonnes versus an official figure of 1,439,910.2 tonnes,

equivalent to a Regency Accuracy of 97.50%. These findings confirm that the satellite-based model can support estimation of rice productivity and production in Karawang Regency.

Keyword : *Rice Productivity Prediction, Sentinel Satellite Imagery, NDVI Time-Series, Machine Learning, Leave-One-Year-Out, Karawang Regency.*

1. PENDAHULUAN

Padi (*Oryza sativa*) merupakan komoditas kunci bagi ketahanan pangan karena menjadi sumber karbohidrat utama bagi mayoritas penduduk Indonesia. Secara operasional, produksi padi suatu wilayah dihitung melalui pendekatan $area \times yield$, yaitu hasil perkalian antara luas panen dan produktivitas per satuan luas, sejalan dengan definisi resmi yang digunakan Badan Pusat Statistik dalam penghitungan produksi tanaman pangan (Badan Pusat Statistik Kabupaten Karawang, 2024). Namun, komponen produktivitas pada pendekatan ini umumnya hanya dapat diperoleh melalui survei ubinan atau survei lapangan langsung, yang membutuhkan waktu, biaya, dan tenaga tidak sedikit untuk mencakup seluruh 30 kecamatan di Kabupaten Karawang secara berkala setiap musim tanam.

Dalam kerangka ketahanan pangan nasional, beras dipandang sebagai komoditas strategis yang belum dapat digantikan dalam waktu dekat (Clapp et al., 2022), sementara Indonesia termasuk produsen sekaligus konsumen beras terbesar di dunia (Food & of the United Nations, 2023). Kabupaten Karawang secara historis dikenal sebagai salah satu lumbung padi utama di Provinsi Jawa Barat. Namun, alih fungsi lahan akibat pesatnya perkembangan kawasan industri, dinamika perubahan iklim, serta keterbatasan pemantauan berbasis survei lapangan membuat kebutuhan akan prediksi produksi padi yang akurat sekaligus tepat waktu pada skala kabupaten menjadi semakin mendesak (Wu et al., 2023).

Penginderaan jauh menyediakan alternatif observasi yang objektif, berulang, dan terukur baik secara spasial maupun temporal (Richards, 2022). Citra Sentinel-2 menawarkan data multispektral beresolusi 10 meter dengan pita Merah dan *Near-Infrared* (NIR) yang memungkinkan perhitungan *Normalized Difference Vegetation Index* (NDVI). Sejumlah kajian menunjukkan bahwa NDVI mampu merepresentasikan perubahan fenologi padi, sehingga deret waktu NDVI dapat digunakan sebagai indikator produktivitas (Aslan et al., 2024; European Space Agency, 2023). Berbagai penelitian sebelumnya telah memanfaatkan citra satelit dan *machine learning* untuk memprediksi produktivitas maupun memetakan luas tanam padi, termasuk fusi Sentinel-1 dan Sentinel-2 (Xiao et al., 2021) maupun kombinasi parameter biofisik dengan data SAR dan optik (Sah et al., 2024). Namun, sebagian besar penelitian tersebut masih berfokus pada pemetaan luas tanam atau estimasi produktivitas pada skala plot, termasuk kajian lokal di Kabupaten Karawang yang mengandalkan ambang NDVI statis berbasis indeks vegetasi tunggal tanpa validasi lintas tahun (Supriatna et al., 2020), dan belum mengintegrasikan hasil prediksi produktivitas dengan data luas panen resmi ke dalam kerangka $area \times yield$ pada skala kabupaten menggunakan validasi temporal yang ketat seperti *Leave One Year Out* (LOYO).

Berdasarkan kesenjangan tersebut, penelitian ini menggunakan data periode tahun 2019 sampai 2025 dengan mengintegrasikan fitur deret waktu multisensor mencakup Sentinel-2, Sentinel-1 SAR, dan variabel iklim dari

CHIRPS serta MODIS ke dalam kerangka *machine learning*, dievaluasi menggunakan skema validasi LOYO yang ketat. Tujuan penelitian ini adalah: (1) menerapkan mask sawah berbasis dinamika NDVI dari citra Sentinel-2 untuk mengenali piksel sawah tingkat kecamatan di Kabupaten Karawang; (2) menganalisis pola deret waktu NDVI dalam mencerminkan fase pertumbuhan padi serta menyusun indikator fenologi sebagai fitur pendukung; dan (3) membangun serta mengevaluasi model prediksi produktivitas padi berbasis *machine learning*, sekaligus mengestimasi produksi padi Kabupaten Karawang tahun 2025 melalui pendekatan $area \times yield$ dan membandingkannya dengan data aktual Dinas Pertanian Kabupaten Karawang.

2. LANDASAN TEORI

Penginderaan Jauh Untuk Monitoring Pertanian

Panjang Penginderaan jauh (*remote sensing*) merupakan metode untuk memperoleh informasi mengenai objek melalui perekaman sensor tanpa kontak langsung dengan target, kemudian hasil perekaman tersebut dianalisis guna menafsirkan karakteristiknya (Richards, 2022). Dalam konteks pertanian, pendekatan ini banyak digunakan untuk memantau kondisi vegetasi secara periodik pada beragam skala spasial dan temporal, termasuk pemetaan tutupan lahan, penilaian kesehatan tanaman, serta pengamatan dinamika pertumbuhan yang diperlukan untuk mendukung perencanaan produksi (Saha et al., 2025). Karena mampu mencakup wilayah yang luas dan memiliki frekuensi akuisisi yang relatif teratur, penginderaan jauh menghasilkan bukti kuantitatif yang memungkinkan perbandingan antar musim maupun antar wilayah, sehingga menjadi fondasi berbagai pendekatan di literatur, baik untuk estimasi luasan

panen, pemantauan perubahan pertumbuhan tanaman, maupun perumusan proksi produktivitas suatu wilayah.

Citra Satelit Sentinel-2 dan Sentinel-1 (SAR)

Sentinel-2 adalah konstelasi satelit observasi darat dalam Program Copernicus yang dioperasikan oleh European Space Agency, membawa instrumen *MultiSpectral Instrument* (MSI) yang merekam 13 pita spektral dengan resolusi spasial 10 m, 20 m, dan 60 m serta kunjungan ulang efektif sekitar 5 hari, sehingga mampu menyediakan deret waktu yang memadai untuk memantau dinamika vegetasi (European Space Agency, 2023). Pita Merah dan NIR pada resolusi 10 m menjadi komponen utama pembentuk indeks vegetasi, sementara sebagai sensor optik pasif, Sentinel-2 tetap memiliki keterbatasan berupa gangguan awan dan variasi kondisi atmosfer.

Sentinel-1 merupakan konstelasi satelit *Synthetic Aperture Radar* (SAR) yang juga dioperasikan oleh European Space Agency, memancarkan gelombang mikro aktif pada pita C dan merekam sinyal *backscatter* yang dipantulkan dari permukaan bumi. Keunggulan mendasar sensor aktif ini adalah kemampuannya beroperasi tanpa hambatan awan, sehingga sangat relevan untuk pemantauan wilayah tropis berawan tinggi seperti Kabupaten Karawang (European Space Agency, 2023). Polarisasi VH khususnya responsif terhadap volume dan struktur tajuk tanaman, sehingga deret waktu *backscatter*-nya dapat difungsikan sebagai pelengkap sinyal fenologi optis, terutama pada bulan dengan tutupan awan tinggi (Ginting et al., 2025).

Indeks Vegetasi NDVI dan LSWI

Normalized Difference Vegetation Index (NDVI) merupakan indeks vegetasi yang paling luas digunakan karena formulanya sederhana dan efektif untuk

menelusuri dinamika kehijauan kanopi (Aslan et al., 2024). NDVI dihitung dari reflektansi permukaan pada pita NIR dan Merah dengan persamaan:

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}}$$

dengan ρ_{NIR} berasal dari Band 8 dan ρ_{Red} dari Band 4 pada Sentinel-2. Nilai NDVI berkisar antara -1 hingga +1; nilai mendekati +1 menunjukkan vegetasi hijau lebat. Pada tanaman padi, pola musiman NDVI yang meningkat pada fase vegetatif dan menurun pada fase pemasakan dimanfaatkan untuk penandaan fenologi kanopi (Supriatna et al., 2020). Dari kurva musiman NDVI, berbagai fitur fenologi dapat diekstraksi, seperti nilai puncak (NDVI_{max}), waktu menuju puncak, amplitudo tahunan, serta laju *green-up* dan *senescence* (Meng et al., 2024).

Sebagai pelengkap NDVI, *Land Surface Water Index* (LSWI) memanfaatkan kombinasi pita NIR dan SWIR untuk mendeteksi kandungan air pada jaringan daun maupun permukaan lahan. Pada padi sawah, LSWI sangat informatif selama fase penggenangan dan persiapan lahan, karena pita SWIR peka terhadap kelembapan permukaan yang tidak sepenuhnya tertangkap oleh indeks berbasis NIR–Merah semata (Declaro et al., 2025).

Data Iklim Pendukung

Selain kondisi kanopi yang terekam melalui citra optik dan radar, kondisi iklim turut berperan penting dalam menentukan pertumbuhan dan produktivitas tanaman padi, karena memengaruhi proses fotosintesis, ketersediaan air, dan perkembangan fenologi tanaman sepanjang musim tanam. Curah hujan harian bersumber dari CHIRPS (*Climate Hazards Group InfraRed Precipitation with Station data*), produk estimasi hujan yang keandalannya untuk konteks Indonesia telah divalidasi terhadap puluhan stasiun hujan (Marzuki et al., 2025). Suhu permukaan lahan dan

evapotranspirasi diperoleh dari produk MODIS, yang masing-masing mencerminkan tekanan termal pada tanaman dan keseimbangan air lahan pertanian (Endsley et al., 2025). Ketiga variabel iklim ini lazim disertakan sebagai variabel pendukung dalam studi prediksi hasil pertanian berbasis penginderaan jauh, guna melengkapi sinyal biofisik yang tidak sepenuhnya tertangkap oleh indeks vegetasi maupun data SAR semata.

Produksi dan Produktivitas Padi

Produksi padi merupakan indikator penting dalam evaluasi sektor pertanian karena mencerminkan kemampuan suatu wilayah dalam memenuhi kebutuhan pangan (Badan Pusat Statistik Kabupaten Karawang, 2024). Besarnya produksi ditentukan oleh dua komponen utama, yaitu luas panen dan produktivitas. Salah satu kerangka yang umum digunakan untuk mengestimasi produksi dari kedua komponen tersebut adalah pendekatan $area \times yield$, yang memisahkan proses estimasi luasan dan estimasi hasil per satuan luas sebelum digabungkan menjadi angka produksi akhir (Islam et al., 2023). Pendekatan ini banyak diadopsi pada studi estimasi produksi berbasis penginderaan jauh karena memungkinkan komponen produktivitas diestimasi dari citra satelit, sementara komponen luas panen tetap bersumber dari data resmi instansi terkait, sehingga hasil akhirnya tetap dapat dipertanggungjawabkan secara administratif (Choudhary et al., 2022; Tiwari et al., 2024).

3. METODOLOGI

Metodologi Penelitian ini menetapkan padi di Kabupaten Karawang, Jawa Barat, sebagai objek kajian, mencakup 30 kecamatan pada koordinat WGS84 (EPSG:4326). Data yang digunakan mencakup periode tahun 2019 hingga 2025, dengan total 210 observasi tahunan (30 kecamatan $\times$ 7 tahun). Sumber data meliputi citra

Sentinel-2 Level-2A (NDVI dan LSWI), Sentinel-1 SAR (backscatter VV/VH), data iklim CHIRPS dan MODIS (curah hujan, suhu permukaan lahan, evapotranspirasi), seluruhnya diakses melalui Google Earth Engine, serta data produktivitas dan luas panen resmi dari Dinas Pertanian Kabupaten Karawang sebagai variabel target.

Penelitian mengadopsi kerangka *Cross-Industry Standard Process for Data Mining* (CRISP-DM) melalui enam tahap: pemahaman masalah, pemahaman data, persiapan data, pemodelan, evaluasi, serta penerapan dan estimasi. Tahap pra-pemrosesan citra mencakup penyaringan awan pada *Scene Classification Layer*, penskalaan reflektansi, penyesuaian resolusi antar pita, pemotongan sesuai batas administrasi, dan pembentukan komposit temporal median dekadal. Piksel diklasifikasikan sebagai sawah apabila kurva NDVI musiman menunjukkan pola khas budi daya padi (kenaikan tajam pada fase vegetatif, puncak, lalu penurunan pada fase pemasakan). Fitur pada piksel sawah kemudian diagregasi menjadi statistik tingkat kecamatan per tahun.

Fitur pemodelan dikelompokkan menjadi dua skema: Mode A (murni fitur citra satelit, 51 fitur mencakup statistik NDVI, LSWI, indeks vegetasi pendukung EVI/NDRE/NDWI, statistik backscatter SAR, dan estimasi luas fase tanam) dan Mode B (Mode A ditambah variabel iklim dan fitur historis kecamatan seperti produktivitas tahun sebelumnya dan *cropping intensity*, total 57 fitur). Seleksi fitur menggunakan *permutation importance* pada setiap fold validasi. Lima algoritma diuji secara komparatif: Random Forest ($n_estimators=200$, $max_depth=8$), XGBoost ($n_estimators=200$, $max_depth=4$), LightGBM ($n_estimators=200$, $max_depth=4$, $min_child_samples=5$), Ridge Regression ($alpha=1.0$), dan Ordinary Least Squares, seluruhnya dengan $random_state=42$ untuk menjamin

keterulangan hasil. Pra-pemrosesan citra dilaksanakan pada Google Earth Engine, sedangkan pemodelan dilakukan pada Google Colab menggunakan Python (scikit-learn, XGBoost, LightGBM).

Validasi dilakukan menggunakan dua skema secara paralel: *Leave One Year Out* (LOYO) sebagai skema primer, di mana satu tahun dijadikan data uji secara bergiliran sementara enam tahun lainnya menjadi data latih (7 fold), dan *time split* 80:20 sebagai skema pembandingan independen, dengan data tahun 2019–2023 sebagai data latih dan 2024–2025 sebagai data uji. Performa dinilai menggunakan empat metrik: MAPE sebagai kriteria utama pemilihan model, RMSE, R^2 sebagai metrik pelengkap, dan Akurasi Kabupaten sebagai ukuran kesesuaian estimasi produksi total tingkat kabupaten terhadap data aktual. Model terpilih selanjutnya diterapkan pada besaran operasional melalui pendekatan $area \times yield$, di mana komponen produktivitas diperoleh dari hasil prediksi model, sedangkan komponen luas panen mengacu pada data resmi Dinas Pertanian Kabupaten Karawang.

Tabel 1. Rancangan Penelitian

Komponen	Keterangan
Lokasi	Kabupaten Karawang (30 kecamatan)
Periode data	2019–2025
Unit analisis	Kecamatan
Jumlah observasi	210 (30 kecamatan $\times$ 7 tahun)
Variabel target (Y)	Produktivitas padi (ton/ha)
Variabel masukan (X)	Fitur NDVI, LSWI, SAR (VV/VH), iklim
Pendekatan estimasi produksi	Area $\times$ yield

4. HASIL DAN PEMBAHASAN

Hasil Rata-rata produktivitas padi di Kabupaten Karawang pada periode 2019–2025 tercatat sebesar 6,780 ton/ha (simpangan baku 0,592; rentang 4,704–8,585 ton/ha), dengan

koefisien variasi 8,73%, konsisten dengan produktivitas padi sawah irigasi di Pulau Jawa pada umumnya. Eksplorasi pola musiman NDVI memperlihatkan dua puncak kehijauan kanopi sepanjang tahun: puncak pertama sekitar bulan April yang berasosiasi dengan musim tanam pertama (MT1), dan puncak kedua sekitar bulan Oktober yang berasosiasi dengan musim tanam kedua (MT2). Pola dwimusim ini konsisten pada seluruh tahun pengamatan. Pemeriksaan lebih lanjut pada tahun 2023 menemukan penurunan curah hujan tajam pada Agustus–Oktober yang diikuti penurunan NDVI, konsisten dengan fenomena El Niño yang menyebabkan kekeringan di sebagian besar wilayah Indonesia pada periode tersebut.

Perbandingan skema fitur Mode A dan Mode B pada validasi LOYO menunjukkan bahwa seluruh lima algoritma menghasilkan performa yang lebih baik pada Mode A tanpa kecuali, dengan penurunan performa paling tajam pada model linear (MAPE OLS naik dari 7,285% menjadi 9,249%; Ridge dari 7,675% menjadi 9,343%). Temuan ini mengindikasikan bahwa penambahan fitur iklim dan historis pada Mode B justru memperkenalkan multikolinearitas dan derau tambahan pada jumlah observasi yang relatif terbatas (210 baris). Berdasarkan hasil ini, Mode A ditetapkan sebagai skema fitur utama untuk seluruh tahapan pemodelan selanjutnya.

Tabel 2. Hasil Validasi Leave One Year Out (LOYO), Mode A

Model	MAPE (%)	MSE (ton/ha)	R ²	Akurasi Kabupaten (%)
Random Forest	7,207	,580	−2,358	94,154
OLS	7,285	,585	−2,668	94,721
XGBoost	7,317	,588	−2,400	94,544

LightGBM	7,342	,586	−2,548	94,924
Ridge Regression	7,675	,623	−2,726	93,811

Random Forest memperoleh MAPE dan RMSE terendah pada skema LOYO sehingga dipilih sebagai model terbaik pada tingkat kecamatan, sementara LightGBM menghasilkan Akurasi Kabupaten tertinggi (94,924%) meskipun MAPE-nya sedikit lebih tinggi. Sebagai skema pembandingan independen, validasi *time split* 80:20 (Tabel 3) justru menunjukkan pola berbeda, dengan OLS mencatatkan MAPE terendah.

Tabel 3. Hasil Validasi Time Split 80:20, Mode A

Model	MAPE (%)	MSE (ton/ha)	R ²	Akurasi Kabupaten (%)
OLS	5,388	,480	−0,968	98,509
Random Forest	5,746	,538	−1,478	97,452
Ridge Regression	6,093	,599	−2,074	95,886
LightGBM	6,453	,606	−2,142	97,225
XGBoost	7,010	,605	−2,135	97,951

Mengingat LOYO ditetapkan sebagai skema pelaporan primer karena mensimulasikan kondisi penerapan pada tahun yang sepenuhnya tidak terlihat, penelitian ini tetap menjadikan Random Forest sebagai model terpilih untuk tahap estimasi produksi. Nilai R² negatif pada Tabel 2 dan 3 merupakan artefak statistik akibat rentang produktivitas data uji yang sempit (koefisien variasi 8,73%), bukan indikasi model yang tidak layak, sebagaimana ditunjukkan oleh nilai MAPE yang tetap rendah dan Akurasi Kabupaten yang tetap tinggi. Pada skala produksi absolut (ton) tingkat kecamatan, seluruh model

mencatatkan R^2 di atas 0,95, dengan Random Forest dan OLS unggul tipis pada $R^2 = 0,961$; perbedaan ini terjadi karena pada skala produksi absolut, variansi total data didominasi variasi luas panen antarkecamatan yang jauh lebih lebar dibandingkan variansi produktivitas per-kecamatan.

Tabel 4. Produksi Aktual dan Prediksi Kabupaten Karawang per Tahun (LOYO)

Tahun	Produksi Aktual (ton)	Produksi Prediksi (ton)	Akurasi Kabupaten (%)
2019	1.401.690,0	1.456.154,6	96,1
2020	1.370.802,0	1.456.566,0	93,7
2021	1.319.011,3	1.215.214,4	92,1
2022	1.411.085,6	1.373.349,1	97,3
2023	1.452.497,7	1.576.571,6	91,5
2024	1.392.747,7	1.302.788,6	93,5
2025	1.439.910,2	1.475.946,3	97,5

Akurasi Kabupaten berkisar antara 91,5% (2023) hingga 97,5% (2025), dengan rata-rata 94,5% selama tujuh tahun pengamatan. Akurasi terendah terjadi pada tahun 2023, konsisten dengan anomali kekeringan akibat El Niño, di mana model cenderung *overestimate* sebesar 8,5%. Pengujian pembandingan menggunakan luas panen hasil estimasi citra satelit (bukan data resmi) menghasilkan MAPE produksi yang melonjak ke kisaran 95%, menegaskan bahwa data luas panen resmi Dinas Pertanian tetap diperlukan sebagai komponen area dalam kerangka $area \times yield$.

Sejumlah temuan utama dapat dirangkum dari hasil penelitian ini. Pertama, terpilihnya Random Forest sebagai model utama tetap dipertahankan meskipun LightGBM dan OLS unggul pada metrik Akurasi Kabupaten, karena performanya paling konsisten pada MAPE dan RMSE tingkat kecamatan yang merupakan unit analisis spasial utama penelitian.

Kedua, teridentifikasi mekanisme *error cancellation* yang menjelaskan mengapa peringkat model berbeda antara level kecamatan dan kabupaten: Random Forest menghasilkan kesalahan prediksi yang cenderung bersifat sistematis antarkecamatan, sedangkan LightGBM dan OLS menghasilkan kesalahan yang lebih acak arahnya sehingga saling meniadakan ketika diagregasi. Ketiga, estimasi luas panen berbasis citra satelit belum dapat menggantikan peran data resmi Dinas Pertanian pada penelitian ini.

Dibandingkan dengan penelitian terdahulu di lokasi yang sama, yaitu Supriatna et al. (2020) yang memanfaatkan ambang NDVI statis dengan regresi sederhana tanpa validasi lintas tahun, penelitian ini menerapkan pendekatan *machine learning* modern dengan validasi Leave One Year Out yang lebih ketat. Pendekatan *bottom-up* yang diterapkan, yaitu melatih model pada level kecamatan kemudian mengagregasi hasilnya ke level kabupaten, sejalan dengan konsep *Small Area Estimation* yang lazim digunakan di Indonesia untuk estimasi pada wilayah dengan ukuran sampel kecil. Dari sisi implikasi kebijakan, Akurasi Kabupaten yang dihasilkan penelitian ini memiliki relevansi praktis bagi Dinas Pertanian Kabupaten Karawang sebagai alat bantu pemantauan produksi yang lebih efisien dibandingkan mengandalkan survei lapangan secara penuh, sementara estimasi level kecamatan bermanfaat untuk mengidentifikasi wilayah dengan potensi penyimpangan produksi yang memerlukan verifikasi lapangan lebih lanjut. Metode ini juga berpotensi direplikasi ke kabupaten lain di Jawa Barat maupun wilayah lain di Indonesia dengan karakteristik geografis dan pola tanam padi yang serupa.

5. KESIMPULAN

Penelitian ini berhasil menerapkan mask sawah berbasis dinamika NDVI dari citra Sentinel-2 pada seluruh 30 kecamatan di Kabupaten Karawang sebagai tahap praproses fitur spasial. Pola deret waktu NDVI menunjukkan karakteristik dwimusim yang konsisten, dengan puncak kehijauan pertama sekitar bulan April dan puncak kedua sekitar bulan Oktober, sesuai dengan dua musim tanam utama. Random Forest terbukti menjadi algoritma paling unggul pada level kecamatan dengan MAPE sebesar 7,207% dan Akurasi Kabupaten 94,154% pada skema validasi Leave One Year Out (LOYO), meskipun pada level kabupaten LightGBM dan OLS tampil sedikit lebih unggul akibat mekanisme *error cancellation*. Skema *time split* 80:20 sebagai validasi pembandingan independen menunjukkan hasil yang tidak sepenuhnya konsisten dengan LOYO, dengan OLS mencatatkan MAPE terendah; keputusan pemilihan Random Forest sebagai model utama tetap bertumpu pada keunggulannya di level kecamatan pada skema LOYO sebagai skema validasi primer. Model terpilih digunakan untuk mengestimasi produksi Kabupaten Karawang tahun 2025, menghasilkan prediksi sebesar 1.475.946,3 ton berbanding data aktual 1.439.910,2 ton, setara Akurasi Kabupaten 97,50%. Temuan ini menegaskan bahwa model prediksi berbasis citra satelit multisensor dan *machine learning* dengan validasi LOYO yang ketat dapat difungsikan sebagai instrumen pendukung estimasi produktivitas dan produksi padi pada skala kabupaten, sebagai pelengkap bukan pengganti data resmi luas panen dari Dinas Pertanian.

6. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang dan Dinas Pertanian dan Ketahanan Pangan Kabupaten Karawang atas dukungan data dan fasilitas yang diberikan selama penelitian ini berlangsung.

DAFTAR PUSTAKA

- Aslan, M. F., Sabanci, K., & Aslan, B. (2024). Artificial Intelligence Techniques in Crop Yield Estimation Based on Sentinel-2 Data: A Comprehensive Survey. *Sustainability*, 16(18), 8277. <https://doi.org/10.3390/su16188277>
- Badan Pusat Statistik Kabupaten Karawang. (2024). *Luas Panen, Produktivitas, dan Produksi Padi di Indonesia (Konsep dan Definisi)*.
- Choudhary, A., Gopal, S., & others. (2022). Random Forest for rice yield mapping and prediction using Sentinel-2 data with Google Earth Engine. *Geocarto International*, 37(26), 7692–7716. <https://doi.org/10.1080/10106049.2021.2005024>
- Clapp, J., Moseley, W. G., & Termine, P. (2022). Viewpoint: The case for a six-dimensional food security framework. *Food Policy*, 106, 102164. <https://doi.org/10.1016/j.foodpol.2021.102164>
- Declaro, A., Brown, Z., & Kanae, S. (2025). VAWIlog: A Log-Transformed LSWI–EVI Index for Improved Surface Water Mapping in Agricultural Environments. *Remote Sensing*, 17(16), 2771. <https://doi.org/10.3390/rs17162771>
- Endsley, K. A., Zhao, M., Kimball, J. S., Albrethsen, T., & Devadiga, S. (2025). Improved Global Estimates of Terrestrial Evapotranspiration Using the MODIS and VIIRS Sensors. *Journal of Hydrometeorology*, 26(6), 817–833. <https://doi.org/10.1175/JHM-D-24-0145.1>
- European Space Agency. (2023). *Sentinel-2 MSI: User Guides and Product Specifications*.
- Food, & of the United Nations, A. O. (2023). *FAOSTAT: Crops and Livestock Products*.

- Ginting, F. I., Rudyanto, Fatchurrachman, Shah, R. M., Soh, N. C., Giap, S. G. E., Fiantis, D., Setiawan, B. I., Schiller, S., Davitt, A., & Minasny, B. (2025). High-resolution maps of rice cropping intensity across Southeast Asia. *Scientific Data*, 12, 1408. <https://doi.org/10.1038/s41597-025-05722-1>
- Islam, M. D., Di, L., Qamer, F. M., Shrestha, S., Guo, L., Lin, L., Mayer, T. J., & Phalke, A. R. (2023). Rapid Rice Yield Estimation Using Integrated Remote Sensing and Meteorological Data and Machine Learning. *Remote Sensing*, 15(9), 2374. <https://doi.org/10.3390/rs15092374>
- Marzuki, M., Ramadhan, R., Yusnaini, H., Vonnisa, M., Muharsyah, R., Nugraha, Y., Sari, A. Y., Erajalita, A., & others. (2025). Long-Term Spatial-Temporal Variability, Trends and Extreme Rainfall Events Over Indonesia Based on 43 Years of CHIRPS Data. *International Journal of Climatology*, 45(14), e70107. <https://doi.org/10.1002/joc.70107>
- Meng, L., Li, Y., Shen, R., Zheng, Y., Pan, B., Yuan, W., Li, J., & Zhuo, L. (2024). Large-scale and high-resolution paddy rice intensity mapping using downscaling and phenology-based algorithms on Google Earth Engine. *International Journal of Applied Earth Observation and Geoinformation*. <https://doi.org/10.1016/j.jag.2024.103725>
- Richards, J. A. (2022). *Remote Sensing Digital Image Analysis* (6th ed.). Springer. <https://doi.org/10.1007/978-3-030-82327-6>
- Sah, S., Haldar, D., Singh, R. N., & Nain, A. S. (2024). Rice yield prediction through integration of biophysical parameters with SAR and optical remote sensing data using machine learning models. *Scientific Reports*, 14, 21674. <https://doi.org/10.1038/s41598-024-72624-4>
- Saha, S., Kucher, O. D., Utkina, A. O., & Rebouh, N. Y. (2025). Precision agriculture for improving crop yield predictions: a literature review. *Frontiers in Agronomy*, 7, 1566201. <https://doi.org/10.3389/fagro.2025.1566201>
- Supriatna, Nurdiana, & Handayani. (2020). Rice productivity estimation by Sentinel-2A imagery in Karawang Regency, West Java, Indonesia. *IOP Conference Series: Earth and Environmental Science*, 500(1), 12023. <https://doi.org/10.1088/1755-1315/500/1/012023>
- Tiwari, V., Thorp, K., Tulbure, M. G., Kamruzzaman, M. G., Krupnik, T. J., Sankarasubramanian, A., & Ardon, M. (2024). Advancing food security: Rice yield estimation framework using time-series satellite data and machine learning. *PLOS ONE*, 19(12), e0309982. <https://doi.org/10.1371/journal.pone.0309982>
- Wu, B., Zhang, M., Zeng, H., Tian, F., Potgieter, A. B., Qin, X., Yan, N., Zhao, Y., Dong, Q., Boken, V., Guo, H., Wu, F., Zhao, H., Deronde, B., Tits, L., & Loupian, E. (2023). Challenges and opportunities in remote sensing-based crop monitoring: a review. *National Science Review*, 10(4), nwac290. <https://doi.org/10.1093/nsr/nwac290>
- Xiao, W., Xu, S., & He, T. (2021). Mapping Paddy Rice with Sentinel-1/2 and Phenology- and Object-Based Algorithm: An Implementation in Hangjiahui Plain in China Using GEE Platform. *Remote Sensing*, 13(5), 990. <https://doi.org/10.3390/rs13050990>

