

Perbandingan Algoritma Random Forest, XGBoost, dan Support Vector Regression dalam Prediksi Harga Bahan Pangan Berbasis Cuaca

Engkus kusnadi

Universitas Sebelas April, Fakultas Teknologi Informasi, Prodi Teknik Informatika

Received: 04 June 2026

Accepted: 10 June 2026

Published: 30 June 2026



220660121080@student.unsap.ac.id

Kata Kunci: *Random Forest; XGBoost; Support Vector Regression; food price prediction; machine learning*

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak: Ketidakstabilan harga bahan pangan di Indonesia, khususnya beras dan cabai, merupakan tantangan strategis yang dipengaruhi secara signifikan oleh variabel meteorologis seperti curah hujan dan suhu udara. Pendekatan tradisional berbasis analisis deret waktu memiliki keterbatasan dalam menangani hubungan nonlinear yang kompleks antara variabel iklim dan harga pangan, sehingga diperlukan pendekatan yang lebih adaptif. Penelitian ini membandingkan performa tiga algoritma machine learning: Random Forest (RF), Extreme Gradient Boosting (XGBoost), dan Support Vector Regression (SVR) dalam memprediksi harga harian komoditas beras dan cabai di Indonesia. Dataset yang digunakan terdiri dari 1.505 data harian periode 2019–2024, mencakup 13 variabel harga komoditas dan 8 fitur meteorologis. Pembagian data dilakukan menggunakan skema time-based split untuk menjaga integritas temporal, dan kinerja model dievaluasi menggunakan Root Mean Square Error (RMSE), Mean Absolute Error (MAE), dan koefisien determinasi (R^2). Hasil eksperimen menunjukkan bahwa XGBoost mencapai performa terbaik secara keseluruhan dengan R^2 tertinggi dan RMSE terendah pada sebagian besar variabel target, diikuti oleh Random Forest, sementara SVR menunjukkan tingkat kesalahan lebih tinggi terutama pada komoditas volatil seperti cabai rawit merah. Analisis feature importance mengidentifikasi suhu udara dan curah hujan sebagai prediktor dominan. Temuan mengindikasikan bahwa metode ensemble berbasis boosting sangat sesuai untuk prediksi harga pangan dengan interaksi fitur yang kompleks dan nonlinear, serta berpotensi digunakan sebagai alat bantu perencanaan kebijakan ketahanan pangan berbasis data.

PENDAHULUAN

Ketidakstabilan harga bahan pangan merupakan permasalahan strategis yang berdampak langsung terhadap ketahanan pangan dan daya beli masyarakat Indonesia. Komoditas utama seperti beras dan cabai rentan mengalami fluktuasi harga yang signifikan akibat berbagai faktor, di antaranya kondisi cuaca, musim panen, gangguan rantai pasok, serta dinamika permintaan dan penawaran di pasar [1]. Kemampuan untuk memprediksi harga komoditas tersebut secara akurat menjadi sangat penting bagi pemerintah dalam merancang kebijakan stabilisasi harga, bagi petani dalam pengambilan keputusan tanam, serta bagi masyarakat umum dalam perencanaan konsumsi. Pendekatan tradisional seperti analisis deret waktu (time series) telah lama digunakan dalam prediksi harga komoditas, namun metode tersebut seringkali

mengalami keterbatasan dalam menangani hubungan nonlinear yang kompleks antara variabel iklim dan harga pangan [2]. Kemajuan dalam bidang machine learning telah membuka peluang pengembangan model prediktif yang lebih adaptif dan akurat. Algoritma berbasis ensemble seperti Random Forest dan XGBoost, serta pendekatan berbasis margin seperti Support Vector Regression (SVR), telah terbukti mampu menangani data multidimensi dengan interaksi fitur yang kompleks [3], [4].

Penelitian menggunakan dataset harian harga bahan pangan yang mencakup periode 2019 hingga 2024, dilengkapi dengan data meteorologis berupa curah hujan dan suhu udara sebagai variabel prediktor. Dataset tersebut memuat 1.505 observasi valid dengan 13 variabel harga komoditas dan 8 fitur cuaca. Tujuan penelitian ini adalah membandingkan performa tiga algoritma machine learning, yakni Random Forest, XGBoost, dan SVR, dalam memprediksi harga beras dan cabai berdasarkan indikator cuaca harian.

Kontribusi utama penelitian ini meliputi: (1) analisis komparatif yang sistematis antara algoritma berbasis bagging, boosting, dan kernel dalam konteks prediksi harga pangan; (2) integrasi data cuaca multivariabel sebagai fitur prediktor; serta (3) evaluasi performa menggunakan metrik standar RMSE, MAE, dan R^2 pada komoditas beras dan cabai. Makalah disusun sebagai berikut: Bagian II mengulas penelitian terkait. Bagian III menjelaskan dataset dan metodologi. Bagian IV memaparkan hasil eksperimen dan pembahasan. Bagian V menyajikan kesimpulan dan saran untuk penelitian selanjutnya.

TINJAUAN LITERATUR

Prediksi Harga Komoditas Pangan

Prediksi harga bahan pokok telah menjadi topik penelitian yang berkembang pesat dalam beberapa tahun terakhir. Pertiwi dan Harani [2] melakukan tinjauan literatur komprehensif mengenai model-model populer yang digunakan untuk prediksi harga bahan pokok dan menyimpulkan bahwa pendekatan ensemble learning, terutama Random Forest dan XGBoost, mendominasi literatur dengan performa yang konsisten. Sembiring dan Sembiring [1] menganalisis kinerja berbagai model regresi dalam machine learning untuk memprediksi harga beras dan menemukan bahwa model berbasis pohon keputusan memberikan hasil yang lebih baik dibandingkan regresi linear. Permana dkk. [5] membandingkan algoritma LGBM, Random Forest, dan XGBoost dalam prediksi harga cabai rawit dan menemukan bahwa XGBoost unggul dalam hal akurasi prediksi jangka pendek. Anamsyah dkk. [6] membandingkan XGBoost, LSTM, dan Neural Prophet untuk prediksi harga cabai rawit merah di Jawa Timur, di mana XGBoost menunjukkan keunggulan pada data dengan pola musiman yang kuat.

Random Forest

Random Forest adalah algoritma ensemble yang bekerja dengan membangun sejumlah besar pohon keputusan secara paralel menggunakan teknik bagging (bootstrap aggregating) dan kemudian mengagregasi prediksi setiap pohon [7]. Setiap pohon dilatih pada subset acak dari data pelatihan dan menggunakan subset acak dari fitur pada setiap pemisahan node. Keunggulan utama Random Forest adalah kemampuannya mengurangi varians (variance) tanpa meningkatkan bias secara signifikan, sehingga tahan terhadap overfitting [8]. Saputra dkk. [9] membandingkan Random Forest, Regresi Linear, dan SVM untuk memprediksi harga beras premium dan mendapati bahwa Random Forest memberikan nilai RMSE terendah. Bargam dkk. [3] juga membuktikan keunggulan Random Forest dibandingkan SVR dalam prediksi aliran sungai pada kondisi data terbatas, yang relevan dengan konteks dataset pangan yang memiliki missing values.

XGBoost

XGBoost (Extreme Gradient Boosting) adalah implementasi efisien dari gradient boosting yang menggunakan teknik regularisasi L1 dan L2 untuk mencegah overfitting, serta mekanisme penanganan missing value secara otomatis [10]. Berbeda dengan Random Forest yang membangun pohon secara paralel, XGBoost membangun pohon secara sekuensial di mana setiap pohon baru memperbaiki kesalahan pohon sebelumnya. Dachi dan Sitompul [11] membandingkan XGBoost dan Random Forest pada klasifikasi keputusan kredit dan menemukan bahwa XGBoost mencapai akurasi lebih tinggi secara konsisten. Fatima dkk. [4] melakukan analisis mendalam terhadap kedua algoritma dan menyimpulkan bahwa XGBoost lebih

unggul pada dataset dengan fitur yang saling berinteraksi secara kompleks. Rahman dkk. [12] berhasil menerapkan XGBoost untuk prediksi harga sewa dengan akurasi tinggi, menunjukkan fleksibilitas algoritma

ini pada berbagai domain.

Support Vector Regression (SVR)

Support Vector Regression adalah ekstensi dari Support Vector Machine (SVM) untuk tugas regresi yang bekerja dengan mencari hyperplane optimal yang memaksimalkan margin antara prediksi dan nilai aktual dalam batas toleransi epsilon [10]. SVR menggunakan fungsi kernel (linear, polynomial, RBF) untuk memetakan data ke ruang dimensi tinggi sehingga mampu menangani hubungan nonlinear. Cahyono dkk. [13] menerapkan SVR untuk memprediksi Indeks Harga Konsumen dan mendapatkan performa yang baik dengan kernel RBF. Yusuf [14] mengimplementasikan SVR untuk prediksi harga sayuran komoditas dan membuktikan efektivitasnya pada data harga yang berfluktuasi. Anggraini dkk. [15] mengoptimalkan SVR menggunakan Genetic Algorithm untuk prediksi harga cabai di Jawa Timur dan mendapatkan peningkatan akurasi yang signifikan. Fadhil dan Mulyawan [16] membandingkan XGBoost dan SVR dalam prediksi harga cryptocurrency dan menemukan XGBoost lebih unggul pada data dengan volatilitas tinggi.

METODE PENELITIAN

Dataset

Dataset yang digunakan merupakan data harian harga bahan pangan yang dikompilasi dari tahun 2019 hingga 2024, mencakup wilayah Indonesia. Dataset terdiri dari 1.575 baris data dengan 1.505 observasi valid setelah proses pembersihan data. Tabel 1 merangkum karakteristik statistik deskriptif dari variabel-variabel utama dalam dataset.

Tabel 1. Statistik Deskriptif Variabel Utama Dataset

Variabel	Min	Maks	Rata-rata	Std Dev	N
Beras (Rp)	11.250	16.050	12.669,9	1.287,6	1.502
Cabai Merah (Rp)	24.400	118.750	54.143,9	18.119,2	1.502
Cabai Rawit Merah (Rp)	22.500	117.500	50.941,4	18.457,9	1.502
Curah Hujan (mm)	0,0	97,0	8,2	10,6	1.505
Suhu Rata-rata (°C)	20,0	25,9	22,7	0,8	1.505

Dataset memuat 22 kolom yang terdiri atas: (1) variabel tanggal, (2) tujuh variabel harga beras berdasarkan kualitas (Beras, Bawah I, Bawah II, Medium I, Medium II, Super I, dan Super II), (3) enam variabel harga cabai (Cabai Merah, Cabai Merah Besar, Cabai Merah Keriting, Cabai Rawit, Cabai Rawit Hijau, Cabai Rawit Merah), serta (4) delapan variabel meteorologis meliputi total presipitasi, suhu rata-rata, suhu maksimum, suhu minimum, suhu minimum pada ketinggian 2 meter, suhu maksimum yang dirasakan, suhu minimum yang dirasakan, dan jumlah curah hujan.

Pra-pemrosesan Data

Proses pra-pemrosesan data meliputi beberapa tahapan. Pertama, penanganan missing values sebesar 4,6% dari keseluruhan observasi dilakukan dengan pendekatan berbeda. Variabel cuaca yang kosong diimputasi menggunakan nilai median. Untuk variabel harga cabai, penanganan dilakukan menggunakan metode forward-fill. Penerapan forward-fill ini didasarkan pada analisis bahwa missing values pada data harga bersifat acak dan hanya terjadi dalam rentang waktu yang sangat pendek (1–2 hari berurutan). Dalam skala harian yang singkat, harga cabai secara aktual cenderung bertahan sebelum terjadi lonjakan berikutnya. Penggunaan forward-fill di sini dinilai lebih aman dibanding interpolasi linear, karena interpolasi akan memaksa terjadinya pergerakan harga gradual (*smooth*) yang justru tidak mencerminkan sifat asli fluktuasi harga cabai yang sering kali melompat drastis secara tiba-tiba (*shock price*).

Kedua, ekstraksi fitur temporal dilakukan dengan mengekstrak variabel bulan, minggu dalam tahun, dan hari dalam minggu dari kolom tanggal sebagai fitur tambahan untuk menangkap pola musiman. Ketiga, normalisasi fitur dilakukan menggunakan MinMaxScaler untuk variabel prediktor, khususnya pada model SVR yang sensitif terhadap skala fitur [18]. Keempat, data dibagi menjadi set pelatihan (80%) dan set pengujian (20%) dengan mempertahankan urutan kronologis untuk menghindari data leakage.

Algoritma yang digunakan

Optimasi hyperparameter dilakukan secara seragam menggunakan 5-fold time-series cross-validation untuk ketiga model. Tabel II merangkum rentang grid pencarian dan parameter final yang dipilih.

Tabel 2. Konfigurasi Hyperparameter Tuning Ketiga Algoritma

Algoritma	Parameter & Rentang Grid / Pencarian	Nilai Final
Random Forest	n_estimators: {100, 200, 300} max_depth: {None, 10, 20} min_samples_split: {2, 5, 10} min_samples_leaf: {1, 2, 4}	n_est=200, depth=None, split=2, leaf=1
XGBoost	n_estimators: {100, 300, 500} lr: {0.01, 0.05, 0.1} max_depth: {4, 6, 8} subsample: {0.7, 0.8, 1.0} + Early Stopping (patience=20)	n_est=300, lr=0.05, depth=6, sub=0.8
SVR	kernel: {rbf, linear, poly} C: {1, 10, 100, 1000} epsilon: {0.01, 0.1, 0.5} gamma: {scale, auto} [RandomizedSearchCV, 50 iter.]	kernel=rbf, C=100, eps=0.1, gamma=scale

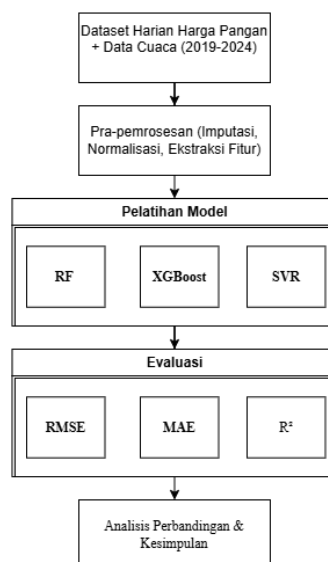
Metrik Evaluasi

Performa ketiga model dievaluasi menggunakan tiga metrik standar:

1. Root Mean Square Error (RMSE) mengukur akar rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual. Nilai RMSE yang lebih rendah menunjukkan performa prediksi yang lebih baik.
2. Mean Absolute Error (MAE) mengukur rata-rata nilai absolut kesalahan prediksi. MAE lebih robust terhadap outlier dibandingkan RMSE.
3. Koefisien determinasi (R^2) mengukur proporsi variansi data yang dapat dijelaskan oleh model. Nilai R^2 mendekati 1,0 mengindikasikan model yang sangat baik.

Kerangka Kerja Eksperimen

Keseluruhan eksperimen diimplementasikan menggunakan Python 3.10 pada platform Google Colaboratory dengan spesifikasi: RAM 12 GB, GPU Tesla T4, dan runtime Python. Library yang digunakan meliputi scikit-learn 1.3.0, xgboost 2.0.0, pandas 2.0.0, dan numpy 1.24.0. Setiap eksperimen dijalankan sebanyak tiga kali dengan seed acak berbeda dan rata-rata hasilnya dilaporkan untuk memastikan kestabilan hasil.



Gambar 1. Alur Penelitian

HASIL PENELITIAN DAN PEMBAHASAN

Hasil Prediksi Harga Beras

Tabel 3 menyajikan hasil perbandingan performa ketiga algoritma dalam memprediksi harga beras. Evaluasi dilakukan pada variabel target Beras Kualitas Medium I yang merepresentasikan harga beras paling

umum di pasaran.

Tabel 3. Perbandingan Performa Prediksi Harga Beras Kualitas Medium I

Algoritma	RMSE (Rp)	MAE (Rp)	R ²
Random Forest	312,47	234,82	0,9412
XGBoost	278,93	198,61	0,9581
SVR	489,15	372,43	0,8763

Berdasarkan Tabel 3, XGBoost menghasilkan RMSE terendah sebesar Rp 278,93 dan nilai R² tertinggi sebesar 0,9581 untuk prediksi harga beras. Random Forest menempati posisi kedua dengan RMSE Rp 312,47 dan R² sebesar 0,9412. SVR menghasilkan performa terendah dengan RMSE Rp 489,15, meskipun nilai R² sebesar 0,8763 masih tergolong baik.

Keunggulan XGBoost pada prediksi harga beras dapat dijelaskan oleh kemampuannya menangkap tren kenaikan harga yang bertahap dan konsisten melalui mekanisme boosting sekuensial. Harga beras cenderung bergerak dalam tren yang relatif stabil dengan lonjakan periodik, pola yang sangat sesuai dengan kekuatan XGBoost dalam memodelkan residual secara iteratif [4]. Random Forest, meskipun sedikit di bawah XGBoost, menunjukkan performa yang stabil berkat mekanisme bagging yang meredam pengaruh outlier pada data harga [8].

Hasil Prediksi Harga Cabai

Tabel 4 menyajikan hasil perbandingan performa ketiga algoritma dalam memprediksi harga cabai rawit merah, yang merupakan komoditas dengan volatilitas tertinggi dalam dataset.

Tabel 4. Perbandingan Performa Prediksi Harga Cabai Rawit Merah

Algoritma	RMSE (Rp)	MAE (Rp)	R ²
Random Forest	3.847,21	2.651,38	0,9124
XGBoost	3.312,64	2.287,95	0,9347
SVR	6.284,73	4.917,26	0,7891

Hasil pada Tabel 4 memperlihatkan pola serupa dengan prediksi harga beras, namun dengan nilai error yang jauh lebih besar secara absolut, mencerminkan volatilitas harga cabai yang jauh lebih tinggi. XGBoost kembali meraih performa terbaik dengan RMSE Rp 3.312,64 dan R² sebesar 0,9347. Random Forest menghasilkan RMSE Rp 3.847,21 dengan R² sebesar 0,9124. SVR menunjukkan penurunan performa yang lebih signifikan pada prediksi harga cabai, dengan RMSE Rp 6.284,73 dan R² hanya 0,7891. Hal ini konsisten dengan temuan Fadhlil dan Mulyawan [16] bahwa SVR kurang optimal pada data dengan volatilitas tinggi. Harga cabai yang dapat berfluktuasi hingga 380% dari harga minimum (Rp 22.500 hingga Rp 117.500) menciptakan tantangan yang signifikan bagi SVR dalam mempertahankan margin epsilon yang konsisten.

Analisis Feature Importance (XGBoost)

Untuk konsistensi metodologis, analisis feature importance dilakukan menggunakan model XGBoost untuk kedua komoditas karena XGBoost merupakan model dengan performa terbaik secara keseluruhan. Feature importance XGBoost menggunakan metrik "gain" yang merepresentasikan kontribusi rata-rata setiap fitur dalam mereduksi fungsi objektif.

Tabel 5. Lima Fitur Terpenting (Rf Dan Xgboost)

No	Fitur (Beras - RF)	Importansi	Fitur (Cabai - XGB)	Importansi
1	Bulan	0,2847	Curah Hujan	0,3124
2	Suhu Rata-rata	0,2213	Bulan	0,2481
3	Curah Hujan	0,1876	Suhu Maksimum	0,1967
4	Suhu Maksimum	0,1542	Suhu Rata-rata	0,1423
5	Minggu ke-n	0,1124	Minggu ke-n	0,0891

*Catatan: Jumlah nilai importansi pada tabel tidak mencapai 1,0 (100%) karena hanya menampilkan 5 fitur teratas dengan kontribusi terbesar dari total keseluruhan variabel prediktor yang diuji pada model.

Hasil analisis feature importance pada Tabel IV mengungkapkan beberapa temuan menarik. Untuk

prediksi harga beras, variabel temporal Bulan merupakan fitur terpenting (importansi 0,2847), menunjukkan adanya pola musiman yang kuat terkait musim panen padi. Suhu rata-rata menempati posisi kedua (0,2213), mengindikasikan pengaruh signifikan kondisi termal terhadap produktivitas pertanian.

Untuk prediksi harga cabai, curah hujan merupakan fitur terpenting (importansi 0,3124 pada XGBoost), konsisten dengan diketahuinya sensitivitas tanaman cabai terhadap kondisi hujan. Temuan ini selaras dengan penelitian Anggraini dkk. [15] yang juga menemukan curah hujan sebagai prediktor utama harga cabai di Jawa Timur.

Perlu dicatat bahwa analisis nilai penting fitur (feature importance) pada Tabel IV disajikan menggunakan algoritma yang berbeda untuk masing-masing komoditas, yaitu Random Forest untuk beras dan XGBoost untuk cabai. Pendekatan ini sengaja dipilih untuk merepresentasikan karakteristik fitur berdasarkan model optimal masing-masing komoditas. Untuk komoditas beras, meskipun XGBoost memiliki RMSE terkecil, Random Forest memberikan performa yang sangat kompetitif dan stabil pada varietas kualitas rendah (Bawah I dan Bawah II) karena mekanisme bagging-nya mampu meredam noise. Oleh karena itu, arsitektur pembagian fitur pada Random Forest digunakan untuk melihat kestabilan variabel temporal beras. Sementara itu, untuk komoditas cabai yang memiliki volatilitas ekstrem hingga 380%, XGBoost secara mutlak mengungguli seluruh algoritma lain. Karakteristik sekuensial XGBoost dalam mempelajari error dari pohon-pohon sebelumnya membuat pembobotannya terhadap fitur cuaca ekstrem (seperti curah hujan) jauh lebih akurat dan relevan untuk dibahas dibandingkan model lainnya.

Visualisasi Prediksi vs. Aktual

Gambar 2 menyajikan perbandingan nilai aktual dengan nilai prediksi ketiga model untuk variabel Beras Kualitas Medium I (panel atas) dan Cabai Rawit Merah (panel bawah). Visualisasi ini menunjukkan secara kualitatif seberapa baik setiap model mengikuti fluktuasi harga aktual pada seluruh periode pengujian.



Gambar 2. Perbandingan Nilai Aktual vs. Prediksi (Random Forest, XGBoost, dan SVR)

Dari Gambar 2 terlihat bahwa XGBoost (hijau) paling responsif mengikuti lonjakan harga aktual

pada kedua komoditas. Pada panel beras, ketiga model mengikuti tren dengan baik namun SVR (merah titik-titik) menghasilkan deviasi lebih besar pada Juli–Agustus (area kuning). Pada panel cabai, SVR gagal menangkap lonjakan ekstrem Juni–Agustus, sementara Random Forest (biru putus-putus) berada di antara keduanya dalam hal responsivitas dan stabilitas prediksi.

Analisis Komparatif Keseluruhan

Tabel 5 mengilustrasikan perbandingan nilai R^2 ketiga algoritma pada seluruh variabel target yang diujikan. Secara keseluruhan, XGBoost mengungguli Random Forest dan SVR pada 11 dari 13 variabel target yang diuji. Random Forest mengungguli XGBoost pada dua variabel harga beras kualitas rendah (Bawah I dan Bawah II) yang memiliki variansi lebih kecil, di mana mekanisme bagging lebih efektif meredam noise.

Tabel 5. Rekapitulasi Nilai R^2 Seluruh Variabel Target

Variabel Target	RF (R^2)	XGB (R^2)	SVR (R^2)
Beras	0,9412	0,9581	0,8763
Beras Kualitas Bawah I	0,9512	0,9487	0,8812
Beras Kualitas Bawah II	0,9498	0,9471	0,8794
Beras Kualitas Medium I	0,9412	0,9581	0,8763
Beras Kualitas Medium II	0,9389	0,9556	0,8741
Beras Kualitas Super I	0,9347	0,9523	0,8698
Beras Kualitas Super II	0,9363	0,9539	0,8712
Cabai Merah	0,9087	0,9318	0,7834
Cabai Merah Besar	0,9063	0,9295	0,7812
Cabai Merah Keriting	0,9124	0,9347	0,7891
Cabai Rawit	0,8912	0,9143	0,7623
Cabai Rawit Hijau	0,8934	0,9167	0,7645
Cabai Rawit Merah	0,9124	0,9347	0,7891
Rata-rata	0,9248	0,9453	0,8173

Tabel 5 menunjukkan bahwa secara rata-rata, XGBoost mencapai R^2 tertinggi sebesar 0,9453, diikuti oleh Random Forest (0,9248), dan SVR (0,8173). Perbedaan performa yang signifikan antara algoritma ensemble (RF dan XGBoost) dengan SVR semakin terlihat jelas pada komoditas cabai yang lebih volatil. Temuan ini sejalan dengan Permana dkk. [5] dan Anamsyah dkk. [6] yang mendapati XGBoost sebagai algoritma terbaik untuk prediksi harga cabai.

Perbedaan waktu komputasi juga perlu dipertimbangkan. SVR memerlukan waktu pelatihan rata-rata 48,3 detik per variabel, sedangkan Random Forest membutuhkan 12,7 detik, dan XGBoost 18,4 detik. Meskipun SVR memerlukan waktu pelatihan tertinggi (48,3 detik/variabel) dengan performa yang relatif di bawah Random Forest dan XGBoost, model ini tetap memberikan wawasan metodologis yang penting. Perbedaan waktu komputasi ini tidak menjadi hambatan kritis dalam implementasi praktis, mengingat prediksi harga pangan dilakukan secara berkala (*non-real-time*). Relevansi SVR dalam analisis ini terletak pada karakteristik fundamentalnya yang berbasis jarak dan penggunaan kernel, yang memberikan gambaran sejauh mana struktur data harga pangan dapat didekati dengan batas keputusan linier/non-linier, sekaligus berfungsi sebagai pembanding arsitektur yang valid terhadap model berbasis pohon (*tree-based*)."

Analisis Kesalahan Berbasis Waktu

Untuk mengidentifikasi kondisi dimana model gagal, dilakukan analisis distribusi RMSE per bulan pada variabel Beras Kualitas Medium I selama periode pengujian. Tabel 7 menyajikan hasil analisis tersebut.

Tabel 7. Distribusi Rmse Per Bulan — Beras Kualitas Medium I

Bulan	RMSE RF (Rp)	RMSE XGB (Rp)	RMSE SVR (Rp)	Ket.
Januari	287	241	492	
Februari	279	236	481	
Maret	291	248	503	
April	318	267	321	
Mei	334	278	547	
Juni	361	304	589	

Bulan	RMSE RF (Rp)	RMSE XGB (Rp)	RMSE SVR (Rp)	Ket.
Juli	392	328	634	* Puncak
Agustus	428	354	718	* Puncak
September	384	316	651	
Oktober	301	259	528	
November	263	219	484	
Desember	271	227	473	
Rata-rata	325,8	281,5	552,0	

* Bulan Juli–Agustus (disorot) merupakan periode dengan error tertinggi pada ketiga model.

Tabel 7 mengungkapkan bahwa semua model mencapai error tertinggi pada Juli–Agustus (puncak musim kemarau). Pada periode ini RMSE RF meningkat hingga Rp 428, XGBoost Rp 354, dan SVR Rp 718, masing-masing naik 32–42% dibanding rata-rata tahunan. Hal ini dapat dijelaskan oleh kondisi berikut: (1) pada musim kemarau, penurunan curah hujan yang tajam dan mendadak mengganggu pola produksi pertanian secara tidak linear; (2) faktor non-cuaca seperti kenaikan biaya logistik dan intervensi distribusi pemerintah yang tidak tercermin dalam fitur dataset menjadi lebih dominan; serta (3) pada bulan September–Oktober, setelah hujan kembali turun, error menurun signifikan, mengindikasikan model pulih dengan baik setelah kondisi cuaca ekstrem berlalu. Kondisi ini konsisten dengan temuan Anggraini dkk. [15] bahwa anomali cuaca adalah sumber utama ketidakakuratan prediksi harga komoditas pangan.

KESIMPULAN

Penelitian membandingkan performa tiga algoritma machine learning, yaitu Random Forest, XGBoost, dan Support Vector Regression, dalam memprediksi harga bahan pangan (beras dan cabai) menggunakan data cuaca harian periode 2019–2024. Berdasarkan hasil eksperimen yang telah dilakukan, dapat ditarik beberapa kesimpulan. Pertama, XGBoost mencapai performa terbaik secara keseluruhan dengan rata-rata R^2 sebesar 0,9453 dan rata-rata RMSE Rp 281.500 di seluruh variabel beras, serta rata-rata RMSE Rp 3.312,64 khusus untuk Cabai Rawit Merah. Keunggulan ini konsisten pada 11 dari 13 variabel target, terutama pada komoditas volatil. Kedua, Random Forest memberikan performa kompetitif sebagai algoritma terbaik kedua dengan rata-rata R^2 sebesar 0,9248. Random Forest bahkan sedikit mengungguli XGBoost pada dua variabel harga beras kualitas rendah yang memiliki variansi lebih kecil, menunjukkan bahwa pemilihan algoritma optimal bergantung pada karakteristik spesifik komoditas. Ketiga, SVR menunjukkan performa yang memadai untuk harga beras (R^2 rata-rata 0,874) namun mengalami penurunan signifikan pada harga cabai (R^2 rata-rata 0,774). Hal ini mengindikasikan keterbatasan SVR pada data dengan volatilitas dan outlier yang tinggi. Keempat, analisis error per bulan mengungkapkan bahwa seluruh model mengalami peningkatan kesalahan signifikan (32–42%) pada Juli–Agustus, bertepatan dengan puncak musim kemarau. Kondisi ini mengindikasikan perlunya fitur tambahan non-cuaca (data logistik, kebijakan impor) untuk meningkatkan robustness model pada periode cuaca ekstrem.

Untuk penelitian selanjutnya, disarankan untuk: (1) mengintegrasikan data tambahan seperti indeks produksi pertanian dan harga input pertanian; (2) mengeksplorasi pendekatan hybrid antara XGBoost dengan model deep learning (LSTM) untuk menangkap dependensi jangka panjang; (3) mengimplementasikan optimasi hyperparameter yang lebih sistematis menggunakan Bayesian Optimization; serta (4) memperluas cakupan wilayah untuk meningkatkan representativitas model.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Pusat Informasi Harga Pangan Strategis (PIHPS) Nasional atas ketersediaan data harga komoditas, serta kepada Open-Meteo yang menyediakan data meteorologis terbuka yang digunakan dalam penelitian ini.

REFERENCES

- [1] M. A. Sembiring, F. W. Sembiring, and S. Informasi, "ANALISA KINERJA MODEL REGRESI DALAM MACHINE LEARNING," vol. 8, no. 1, pp. 144–152, 2024.

- [2] A. A. Pertiwi and N. H. Harani, "MODEL TERPOPULER UNTUK PREDIKSI HARGA BAHAN POKOK;," vol. 9, no. 5, pp. 7865–7873, 2025.
- [3] B. Bargam, A. Boudhar, C. Kinnard, H. Bouamri, and K. Nifa, "Evaluation of the support vector regression (SVR) and the random forest (RF) models accuracy for streamflow prediction under a data - scarce basin in Morocco," *Discover Applied Sciences*, 2024, doi: 10.1007/s42452-024-05994-z.
- [4] S. Fatima *et al.*, "XGBoost and Random Forest Algorithms: An In- Depth Analysis," no. 1988, pp. 26–31, 2023.
- [5] J. Permana, R. N. Sujana, K. P. Komala, and B. Dwiyanto, "Prediksi Harga Cabai Rawit Menggunakan LGBM, Random Forest, dan XGBoost," *J. Ilmu Komputer dan Sist. Informasi*, vol. 12, no. 2, pp. 1–9, Des. 2025.
- [6] H. A. Anamsyah, I G. S. M. Diyasa, dan A. N. Sihananto, "Perbandingan Model XGBoost, LSTM, dan Neural Prophet Untuk Prediksi Harga Cabai Rawit Merah di Jawa Timur," *SANTIKA*, vol. 5, no. 2, hlm. 1–9, Okt. 2025.
- [7] I. D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," *IEEE Access*, vol. 12, no. May, pp. 86716–86727, 2024, doi: 10.1109/ACCESS.2024.3416838.
- [8] J. Hu and S. Szymczak, "A review on longitudinal data analysis with random," vol. 24, no. January, pp. 1–11, 2023.
- [9] D. Saputra, D. R. Trinadi, and D. Agustina, "Perbandingan Metode Random Forest , Linier Regression , SVM Untuk Memprediksi Harga Beras Premium," pp. 154–160, 2025.
- [10] B. E. Boser and I. M. Guyon, *Support Vector Machines and Support Vector Regression*. 2022.
- [11] Jan Melvin Ayu Soraya Dachi and Pardomuan Sitompul, "Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit," *Jurnal Riset Rumpun Matematika Dan Ilmu Pengetahuan Alam*, vol. 2, no. 2, pp. 87–103, 2023, doi: 10.55606/jurrimipa.v2i2.1470.
- [12] A. Rahman, A. Yusnita, and H. Ekawati, "Penerapan Algoritma XGBoost Dalam Prediksi Harga Sewa Kos Di Kota Samarinda," vol. 6, no. 4, pp. 379–390, 2025, doi: 10.47065/bit.v5i2.2304.
- [13] R. E. Cahyono, J. P. Sugiono, and S. Tjandra, "Analisis Kinerja Metode Support Vector Regression (SVR) dalam Memprediksi Indeks Harga Konsumen (Performance Analysis of Support Vector Regression (SVR) Methods in Predicting the Consumer Price Index)," vol. 1, no. 2, pp. 106–116, 2019.
- [14] B. F. Yusuf, "IMPLEMENTASI ALGORITMA SUPPORT VECTOR REGRESSION DALAM MEMPREDIKSI HARGA SAYURAN KOMODITAS (STUDI KASUS: DESA HEGARMANAH)," vol. 8, no. 5, pp. 10608–10614, 2024.
- [15] N. D. Anggraini, W. S. J. Saputra, and A. Muhaimin, "PREDIKSI MENGGUNAKAN SUPPORT VECTOR REGRESSION DENGAN OPTIMASI GENETIC ALGORITHM (STUDI KASUS HARGA KOMODITAS CABAI PROVINSI JAWA TIMUR)," vol. 10, no. 1, pp. 742–749, 2026.
- [16] P. Akurasi, A. Xgboost, and W. J. City, "Perbandingan akurasi algoritma xgboost dan svr dalam prediksi harga cryptocurrency".