

PREDIKSI HARGA KOMODITAS HORTIKULTURA PANGAN STRATEGIS MENGGUNAKAN RANDOM FOREST DENGAN ANALISIS SHAP

Valencia Patrice Gracia Pangaribuan¹, Bayu Sulistiyanto Ipung Sutejo², Gerry Firmansyah³, Arief Ichwani⁴

Program Studi Teknik Informatika, Fakultas Ilmu Komputer
Universitas Esa Unggul^{1,2,3,4}

Jalan Arjuna Utara No 9, Kebon Jeruk, Jakarta Barat
valenpatriceg@student.esaunggul.ac.id¹,
bayu.sulistiyanto@esaunggul.ac.id², gerry@esaunggul.ac.id³,
arief.ichwani@esaunggul.ac.id⁴

Abstrak

Fluktuasi harga komoditas hortikultura pangan strategis dapat memengaruhi stabilitas pangan dan ekonomi masyarakat. Penelitian ini bertujuan memprediksi harga bawang merah, bawang putih, cabai merah, dan cabai rawit di Provinsi Jawa Tengah menggunakan *Random Forest Regression* berdasarkan harga historis dan variabel iklim, serta menginterpretasikan kontribusi variabel iklim menggunakan *Shapley Additive Explanations* (SHAP). Data harga diperoleh dari Pusat Informasi Harga Pangan Strategis (PIHPS) Nasional, sedangkan data iklim diperoleh dari empat stasiun BMKG selama periode 2018–2025. Model menggunakan 16 fitur, meliputi harga historis, temperatur rata-rata (TAVG), kelembapan rata-rata (RH_AVG), curah hujan (RR), fitur rolling, dan variabel kalender. Evaluasi dilakukan menggunakan MAE, RMSE, dan MAPE pada horizon prediksi H+1, H+7, dan H+30. Hasil evaluasi menunjukkan bahwa performa terbaik diperoleh pada prediksi harga bawang putih H+1 dengan MAPE sebesar 1,69%, sedangkan galat tertinggi terjadi pada cabai merah H+30 sebesar 26,72%. Analisis SHAP menunjukkan bahwa TAVG dominan pada kelompok cabai, sedangkan RR dan RH_AVG lebih berpengaruh pada bawang merah dan bawang putih. Temuan ini menunjukkan bahwa *Random Forest* dan SHAP dapat mendukung prediksi harga jangka pendek dengan interpretasi kontribusi variabel iklim.

Kata Kunci : Random Forest Regression, SHAP, prediksi harga, komoditas hortikultura, iklim

Abstract

Price fluctuations in strategic horticultural food commodities can affect food and economic stability. This study aims to predict the prices of shallots, garlic, red chilies, and bird's eye chilies in Central Java Province using Random Forest Regression based on historical prices and climate variables, and to interpret climate variable contributions using Shapley Additive Explanations (SHAP). Price data were obtained from the National Strategic Food Price Information Center (PIHPS), while climate data were collected from four BMKG stations during 2018–2025. The model used 16 features, including historical prices, average temperature (TAVG), average humidity (RH_AVG), rainfall (RR), rolling features, and calendar variables. Evaluation was conducted using MAE, RMSE, and MAPE at H+1, H+7, and H+30 forecasting horizons. The results show that the best performance was achieved for garlic price prediction at H+1, with a MAPE of 1.69%, while the highest error occurred for red chili at H+30, with a MAPE of 26.72%. SHAP analysis shows that TAVG was dominant in chili commodities, while RR and RH_AVG had greater influence on shallots and garlic. These findings indicate that Random Forest and SHAP can support short-term price prediction with interpretable climate variable contributions.

Keyword : Random Forest Regression, SHAP, price forecasting, horticultural commodities, climate

PENDAHULUAN

Sektor pertanian memegang peranan strategis dalam mendukung ketersediaan pangan serta stabilitas ekonomi nasional. Salah satu persoalan yang sering terjadi pada sektor pangan adalah fluktuasi harga komoditas hortikultura pangan strategis, seperti bawang merah, bawang putih, cabai merah, dan cabai rawit. Komoditas tersebut mempunyai karakteristik mudah rusak, bergantung pada musim panen,

serta rentan terhadap gangguan produksi dan distribusi. Fluktuasi harga pada komoditas pangan strategis dapat memengaruhi daya beli masyarakat, kesejahteraan petani, serta inflasi pangan (Ridwansyah Matondang & Krisnamurthi, 2024). Berdasarkan data PIHPS periode 2018–2025, harga beberapa komoditas menunjukkan peningkatan dan perubahan yang cukup dinamis, seperti cabai merah yang meningkat dari sekitar Rp30.084/kg pada tahun 2018 menjadi Rp53.569/kg pada tahun 2025, serta bawang merah dari sekitar Rp24.168/kg menjadi Rp42.342/kg. Kondisi tersebut menunjukkan bahwa pergerakan harga komoditas hortikultura perlu dipantau dan diprediksi secara lebih adaptif menggunakan pendekatan berbasis data.

Pergerakan harga komoditas hortikultura ditentukan oleh banyak hal, di mana pola harga pada periode sebelumnya hanyalah satu dari sejumlah faktor penentu. Selain itu, aspek lingkungan, terutama iklim, ikut memainkan peran yang cukup besar. Temperatur, kelembapan, dan curah hujan dapat memengaruhi pertumbuhan tanaman, produktivitas, kualitas hasil panen, serta kontinuitas pasokan. Penelitian sebelumnya menunjukkan bahwa variabel iklim dan meteorologis dapat berkontribusi terhadap dinamika harga komoditas pertanian karena berhubungan dengan produksi dan ketersediaan pasokan (Kaewchada et al., 2023; Kumar et al., 2025). Atas dasar pertimbangan tersebut, integrasi antara harga historis dan variabel iklim menjadi penting dalam membangun model prediksi harga yang lebih komprehensif. Random Forest Regression dipandang sebagai pendekatan yang sesuai karena cukup andal dalam mengidentifikasi keterkaitan non-linear di antara variabel dan telah terbukti memiliki performa baik dalam prediksi harga komoditas pertanian (Kaewchada et al., 2023; Ibrahim, 2025).

Meskipun Random Forest dapat menghasilkan prediksi yang baik, model pembelajaran mesin sering dianggap sulit diinterpretasikan secara langsung. Oleh karena itu, diperlukan metode interpretabilitas agar hasil prediksi tidak hanya menunjukkan nilai estimasi, tetapi juga dapat menjelaskan kontribusi variabel terhadap hasil prediksi. Shapley Additive Explanations (SHAP) banyak dipakai sebagai alat penjelas pada model berbasis pohon, Random Forest salah satunya (Bifarin, 2023). Dalam konteks prediksi harga komoditas hortikultura, SHAP memungkinkan peneliti melacak parameter iklim mana yang memberi pengaruh paling kuat pada hasil prediksi, seperti yang telah dilakukan Ardian et al. (2024) dalam prediksi harga cabai rawit berbasis iklim harian.

Berdasarkan latar belakang tersebut, penelitian ini berupaya menjawab persoalan mengenai penerapan *Random Forest Regression* dalam memproyeksikan harga empat komoditas hortikultura, yaitu bawang merah, bawang putih, cabai merah, dan cabai rawit di Jawa Tengah, dengan acuan data harga historis dan kondisi iklim, bagaimana performa model pada horizon H+1, H+7, dan H+30, serta variabel iklim apa yang paling berkontribusi terhadap hasil prediksi berdasarkan analisis SHAP. Penelitian ini bertujuan membangun model peramalan harga harian untuk komoditas hortikultura pangan strategis, mengukur tingkat akurasi model melalui indikator MAE, RMSE, dan MAPE, serta menganalisis kontribusi variabel iklim menggunakan SHAP. Luaran dari penelitian ini diharapkan turut memperluas pemanfaatan machine learning pada bidang ketahanan pangan dan pertanian, sekaligus menjadi bahan pertimbangan dalam pemantauan dan antisipasi fluktuasi harga bagi pemerintah, pelaku pasar, dan pemangku kepentingan di sektor pangan.

PENELITIAN RELEVAN

Studi publikasi yang dilakukan oleh Ardian et al. (2024) menggunakan data iklim harian untuk memprediksi harga cabai rawit di Semarang dengan membandingkan model Random Forest, XGBoost, dan KNN, serta menerapkan SHAP untuk menganalisis kontribusi variabel terhadap hasil prediksi. Hasil penelitian tersebut menunjukkan bahwa suhu menjadi salah satu variabel iklim yang berpengaruh terhadap harga cabai rawit. Penelitian ini relevan karena sama-sama menggunakan data iklim harian dan SHAP. Akan tetapi, lingkup kajian pada penelitian ini diperluas hingga mencakup empat jenis komoditas hortikultura pangan strategis.

Ibrahim. (2025) memanfaatkan Random Forest Regression guna meramalkan harga bawang merah mengandalkan rekam jejak harga di periode sebelumnya dan fitur lag dengan hasil MAPE sebesar 10,31%. Lestari et al. (2025) juga menggunakan pendekatan machine learning untuk memprediksi harga bawang merah, bawang putih, cabai merah, dan cabai rawit dengan variabel harga historis, curah hujan, jumlah hari hujan, dan inflasi. Berdasarkan studi tersebut, diperoleh temuan bahwa algoritma Random Forest mencapai kinerja paling optimal saat diterapkan pada komoditas bawang putih, dengan MAPE sebesar 9,07%. Selain itu, Kaewchada et al. (2023) dan (Antad et al., 2024)

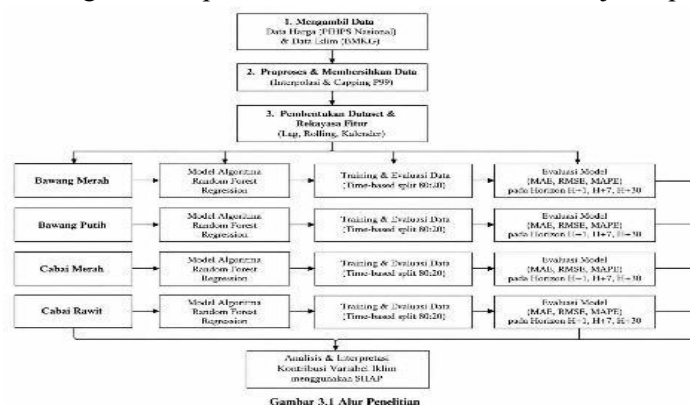
turut membuktikan algoritma Random Forest memperlihatkan hasil evaluasi yang baik dalam prediksi harga komoditas pertanian karena dapat menangkap hubungan non-linear antara harga historis, faktor waktu, dan kondisi cuaca.

Selain aspek prediksi, interpretabilitas model juga menjadi dasar penting dalam penelitian ini. Bifarin (2023) menjelaskan bahwa SHAP dapat digunakan untuk menginterpretasikan model Random Forest dan XGBoost dengan menunjukkan informasi kontribusi fitur secara lokal maupun global. Hal ini penting karena model machine learning sering dianggap sulit dijelaskan secara langsung. Sementara itu, Ridwansyah Matondang & Krisnamurthi (2024) menunjukkan bahwa komoditas pangan strategis seperti bawang merah, bawang putih, dan cabai merah memiliki volatilitas harga yang fluktuatif serta dipengaruhi oleh faktor musiman, cuaca, dan perubahan permintaan pasar. Temuan tersebut memperkuat urgensi penggunaan model prediktif berbasis data dalam menganalisis harga komoditas hortikultura pangan strategis.

Berdasarkan penelitian-penelitian tersebut, dapat disimpulkan bahwa Random Forest telah banyak digunakan dalam prediksi harga komoditas pertanian, sedangkan SHAP dapat digunakan untuk menjelaskan kontribusi variabel dalam model. Namun, beberapa penelitian sebelumnya masih terbatas pada satu komoditas, hanya menggunakan harga historis, atau belum menggabungkan prediksi multi-horizon dengan analisis kontribusi variabel iklim. Oleh karena itu, penelitian ini mengembangkan model prediksi harga harian bawang merah, bawang putih, cabai merah, dan cabai rawit di Provinsi Jawa Tengah menggunakan *Random Forest Regression* berdasarkan harga historis dan variabel iklim, serta menganalisis kontribusi variabel iklim menggunakan SHAP pada horizon H+1, H+7, dan H+30.

METODE PENELITIAN

Studi ini disusun memakai metode kuantitatif dengan berorientasi pada peramalan dengan teknik *machine learning*. Penelitian bertujuan membangun model prediksi harga komoditas hortikultura pangan strategis dengan algoritma *Random Forest Regression* menggunakan harga historis dan variabel iklim. Cakupan studi dibatasi pada peramalan harga harian untuk empat jenis komoditas, meliputi cabai rawit, cabai merah, bawang putih, dan bawang merah, yang seluruhnya berasal dari wilayah Provinsi Jawa Tengah, dengan rentang waktu pengamatan terhitung sejak 1 Januari 2018 hingga 31 Desember 2025. Populasi penelitian mencakup seluruh data harga harian keempat komoditas tersebut dan seluruh data iklim harian yang meliputi temperatur rata-rata (TAVG), kelembapan rata-rata (RH_AVG), dan curah hujan (RR). Sampel penelitian adalah seluruh observasi yang memenuhi kelengkapan data setelah dilakukan penyelarasan tanggal antara data harga dan data iklim. Berdasarkan proses pembentukan fitur dan target, jumlah sampel akhir pada setiap komoditas adalah 2005 observasi untuk horizon H+1, 1999 observasi untuk horizon H+7, dan 1976 observasi untuk horizon H+30. Diagram alur penelitian secara keseluruhan disajikan pada Gambar 1.



Gambar 1. Alur Penelitian

Pengumpulan Data

Data harga harian pada studi ini bersumber dari laman Pusat Informasi Harga Pangan Strategis (PIHPS) Nasional, yang mencakup empat komoditas, yaitu bawang merah, bawang putih, cabai merah, dan cabai rawit. Data iklim harian diperoleh dari Badan Meteorologi, Klimatologi, dan

Geofisika (BMKG). Data iklim yang dimanfaatkan berasal dari empat titik pengamatan di lingkup Jawa Tengah, masing-masing adalah Stasiun Klimatologi Jawa Tengah yang berlokasi di Ungaran, Stasiun Meteorologi Maritim Tegal, Stasiun Meteorologi Tunggul Wulung di Cilacap, serta Stasiun Meteorologi Ahmad Yani di Semarang. Parameter iklim yang dilibatkan dalam analisis terbatas pada tiga indikator, yaitu TAVG, RH_AVG, dan RR.

Praproses Data

Pada data harga, proses yang dilakukan meliputi transformasi format data ke bentuk deret waktu harian, konversi nilai harga ke format numerik, penyesuaian format tanggal, penanganan duplikasi tanggal, serta penghapusan observasi yang tidak memiliki informasi harga. Pada data iklim, proses praproses meliputi penyeragaman format tanggal, pembersihan nilai tidak valid, pengubahan kode khusus seperti 8888 dan 9999 menjadi missing value, serta penanganan nilai hilang menggunakan interpolasi berbasis waktu untuk gap pendek. Data iklim dari empat stasiun kemudian diagregasi menjadi rata-rata harian wilayah penelitian. Khusus variabel curah hujan (RR), dilakukan capping pada persentil ke-99 untuk mengurangi pengaruh nilai ekstrem terhadap proses pembelajaran model.

Pembentukan Dataset dan Rekayasa Fitur

Tahap pembentukan dataset dilakukan dengan menggabungkan data harga dan data iklim berdasarkan tanggal yang sama. Dataset kemudian disusun secara terpisah untuk setiap komoditas agar pola harga masing-masing komoditas dapat dipelajari secara spesifik. Fitur prediktor yang digunakan berjumlah 16 fitur, terdiri atas fitur harga historis, variabel iklim harian, fitur rolling iklim, dan fitur kalender. Fitur harga historis meliputi `harga_lag1`, `harga_lag7`, `harga_lag30`, `harga_roll7`, dan `harga_roll30`. Variabel iklim harian meliputi TAVG, RH_AVG, dan RR. Fitur rolling iklim meliputi `TAVG_roll7`, `TAVG_roll30`, `RH_AVG_roll7`, `RH_AVG_roll30`, `RR_roll7`, dan `RR_roll30`. Fitur kalender meliputi hari dan dayofweek. Target prediksi dibentuk menggunakan teknik shift untuk tiga horizon, yaitu H+1, H+7, dan H+30.

Pemodelan Random Forest Regression

Model peramalan dalam penelitian ini dikembangkan dengan menggunakan algoritma Random Forest Regression. Algoritma ini termasuk metode ensemble yang bekerja dengan membentuk beberapa pohon keputusan, kemudian menggabungkan hasil prediksinya untuk menghasilkan nilai akhir. Pada kasus regresi, hasil akhir ditentukan berdasarkan nilai rata-rata dari seluruh prediksi yang dihasilkan oleh setiap pohon. Metode ini dipilih karena memodelkan hubungan antarvariabel yang tidak bersifat linear dan memiliki performa yang cukup stabil pada data dengan tingkat kompleksitas tinggi (Schonlau & Zou, 2020; Talekar, 2020). Model dibangun secara terpisah untuk setiap kombinasi komoditas dan horizon prediksi, sehingga terdapat model untuk bawang merah, bawang putih, cabai merah, dan cabai rawit pada horizon H+1, H+7, dan H+30. Parameter yang digunakan adalah `n_estimators = 200`, `max_features = "sqrt"`, `min_samples_leaf = 5`, dan `random_state = 42`.

Pelatihan dan Evaluasi Model

Pengujian model dilakukan menggunakan metode time-based split. Dalam proses ini, 80% data pertama digunakan sebagai data latih, sedangkan 20% data terakhir digunakan sebagai data uji. Pembagian data tidak dilakukan secara acak, melainkan mengikuti urutan waktu untuk menjaga karakteristik pola tetap konsisten. Data pelatihan dimanfaatkan untuk menyusun model, sementara data uji digunakan untuk menilai kemampuan model dalam melakukan prediksi pada periode waktu selanjutnya. Evaluasi model dilakukan menggunakan *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE).

Analisis SHAP

Tahap akhir penelitian adalah analisis interpretabilitas model menggunakan Shapley Additive Explanations (SHAP). SHAP digunakan sebagai metode untuk menginformasikan besarnya kontribusi masing – masing fitur terhadap hasil prediksi pada model berbasis pohon, termasuk Random Forest. (Bifarin, 2023). Dalam penelitian ini, nilai SHAP dihitung pada model yang menggunakan seluruh fitur input, tetapi interpretasi difokuskan pada variabel iklim, yaitu TAVG, RH_AVG, dan RR. Analisis ini dilakukan untuk mengetahui variabel iklim yang paling berkontribusi

terhadap prediksi harga pada masing-masing komoditas dan horizon prediksi.

HASIL DAN PEMBAHASAN

Hasil Dataset dan Perancangan Model

Hasil pengolahan data menunjukkan bahwa dataset harga akhir terdiri atas 2036 observasi harian untuk empat komoditas hortikultura pangan strategis, yaitu bawang merah, bawang putih, cabai merah, dan cabai rawit. Data iklim harian yang berasal dari empat stasiun BMKG berjumlah 2922 observasi. Perbedaan jumlah observasi disebabkan oleh adanya ketidaksesuaian tanggal antara data harga dan data iklim, sehingga hanya data yang memiliki kesesuaian waktu yang digunakan dalam proses penggabungan.

Setelah dilakukan proses penyelarasan berdasarkan tanggal, data harga dan iklim digabungkan menjadi satu dataset berbasis deret waktu (time series). Tahap selanjutnya adalah rekayasa fitur yang menghasilkan 16 variabel prediktor, yang terdiri atas harga historis, variabel iklim harian, fitur rolling iklim, serta fitur kalender. Dataset akhir kemudian disusun secara terpisah untuk masing - masing komoditas agar karakteristik masing-masing harga dapat dipelajari secara lebih spesifik oleh model.

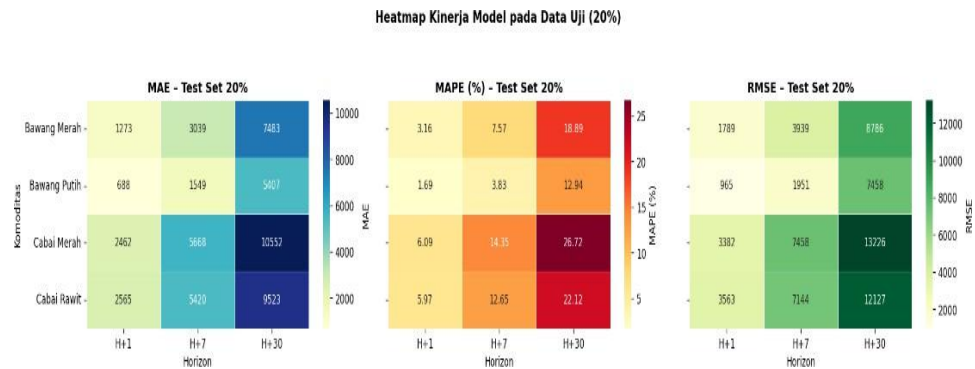
Evaluasi Model Prediksi Harga

Tahap pengujian dilakukan untuk mengetahui sejauh mana model mampu bekerja pada data baru yang belum digunakan dalam tahap pelatihan. Hasil evaluasi model pada data uji disajikan pada Tabel 1. Hasil pengujian diperoleh berdasarkan pembagian data menggunakan metode time-based split, yaitu 80% data latih dan 20% data uji, sehingga urutan waktu tetap terjaga sesuai karakteristik data harga harian.

Tabel 1. Hasil Evaluasi Model pada Data Uji

Komoditas	Horizon	MAE	MAPE (%)	RMSE
Bawang Merah	H+1	1.273	3,16	1.789
Bawang Merah	H+7	3.039	7,57	3.939
Bawang Merah	H+30	7.483	18,89	8.786
Bawang Putih	H+1	688	1,69	965
Bawang Putih	H+7	1.549	3,83	1.951
Bawang Putih	H+30	5.407	12,94	7.458
Cabai Merah	H+1	2.462	6,09	3.382
Cabai Merah	H+7	5.668	14,35	7.458
Cabai Merah	H+30	10.552	26,72	13.226
Cabai Rawit	H+1	2.565	5,97	3.563
Cabai Rawit	H+7	5.420	12,65	7.144
Cabai Rawit	H+30	9.523	22,12	12.127

Berdasarkan Tabel 1, model Random Forest Regression menghasilkan performa terbaik pada horizon H+1 untuk seluruh komoditas. Nilai MAPE terendah diperoleh pada komoditas bawang putih H+1 sebesar 1,69%, dengan MAE sebesar 688 dan RMSE sebesar 965. Sementara itu, galat tertinggi diperoleh pada komoditas cabai merah H+30 dengan MAPE sebesar 26,72%, MAE sebesar 10.552, dan RMSE sebesar 13.226. Hasil penelitian ini menunjukkan bahwa model mampu menghasilkan prediksi dengan tingkat akurasi yang lebih baik pada horizon jangka pendek dibandingkan dengan horizon yang lebih panjang.



Gambar 2. Heatmap Hasil Evaluasi Model pada Data Uji

Berdasarkan Gambar 2, nilai galat meningkat seiring bertambahnya horizon prediksi. Horizon H+1 menghasilkan galat paling rendah, sedangkan H+30 menghasilkan galat paling tinggi pada seluruh komoditas. Hal ini menunjukkan bahwa semakin jauh horizon prediksi, semakin besar ketidakpastian harga akibat pengaruh berbagai faktor eksternal yang tidak sepenuhnya tercakup dalam model. Hasil ini sejalan dengan penelitian Ibrahim (2025) yang menyatakan bahwa pola harga historis memiliki peran penting dalam proses prediksi harga komoditas menggunakan Random Forest Regression.

Apabila ditinjau berdasarkan jenis komoditas, kelompok bawang menunjukkan hasil prediksi yang lebih baik dibandingkan dengan kelompok cabai. Bawang putih menjadi komoditas dengan performa terbaik, sedangkan cabai merah menjadi komoditas dengan tingkat kesalahan tertinggi, terutama pada horizon H+30. Perbedaan ini menunjukkan bahwa karakteristik harga setiap komoditas memengaruhi kemampuan model dalam melakukan prediksi. Komoditas bawang cenderung memiliki pola harga yang lebih stabil, sedangkan cabai merah dan cabai rawit memiliki fluktuasi harga yang lebih tajam. Kondisi tersebut membuat harga cabai lebih sulit diprediksi, khususnya pada horizon yang lebih panjang. Hasil ini konsisten dengan temuan Ridwansyah Matondang & Krisnamurthi (2024) yang menunjukkan bahwa komoditas pangan strategis seperti cabai cenderung memiliki fluktuasi harga yang tinggi.

Keluaran Prediksi Harga

Dalam konteks penelitian ini, keluaran yang dihasilkan berupa model prediksi harga dan tabel hasil prediksi pada data uji. Keluaran prediksi tersebut memuat tanggal, komoditas, horizon prediksi, harga aktual, harga prediksi, residual, error absolut, dan MAPE per observasi. Contoh keluaran prediksi harga bawang merah pada horizon H+1 ditunjukkan pada Tabel 2.

Tabel 2. Contoh Hasil Prediksi Harga Bawang Merah Horizon H+1

Tanggal	Komoditas	Horizon	Harga Aktual	Harga Prediksi	Residual	Error Absolut	MAPE /Obs(%)
18/06/2024	Bawang Merah	H+1	39.900	39.853,21	46,79	46,79	0,12
19/06/2024	Bawang Merah	H+1	39.850	39.465,36	384,64	384,64	0,97
20/06/2024	Bawang Merah	H+1	39.200	39.517,81	-317,81	317,81	0,81
21/06/2024	Bawang Merah	H+1	38.000	39.432,48	-1.432,48	1.432,48	3,77
24/06/2024	Bawang Merah	H+1	37.500	38.731,68	-1.231,68	1.231,68	3,28

Berdasarkan Tabel 2, hasil prediksi harga bawang merah pada horizon H+1 menunjukkan bahwa nilai prediksi model cenderung mendekati harga aktual pada beberapa observasi. Pada tanggal 18 Juni 2024, harga aktual bawang merah sebesar 39.900, sedangkan harga prediksi sebesar 39.853,21 dengan MAPE per observasi sebesar 0,12%. Kondisi tersebut menunjukkan bahwa model cukup mampu menyesuaikan hasil prediksi dengan pola harga aktual, terutama pada prediksi jangka pendek. Namun demikian, pada beberapa periode tertentu masih ditemukan selisih yang lebih besar antara harga aktual dan harga prediksi. Perbedaan tersebut dapat disebabkan oleh fakta bahwa harga harian komoditas hortikultura tidak hanya ditentukan oleh harga historis dan faktor iklim, tetapi juga dipengaruhi oleh faktor eksternal lainnya seperti ketersediaan pasokan, distribusi, permintaan pasar, serta kebijakan yang berlaku.

Secara umum, hasil prediksi pada data pengujian menunjukkan bahwa memiliki kemampuan yang lebih baik dalam menyesuaikan hasil prediksi dengan pola harga aktual pada horizon jangka pendek. Pada horizon yang lebih panjang, hasil prediksi cenderung lebih stabil dan tidak mengikuti fluktuasi harga aktual secara detail. Kondisi ini menunjukkan bahwa *Random Forest Regression* mampu mengenali pola umum yang terdapat dalam data, namun memiliki keterbatasan dalam merepresentasikan perubahan harga yang bersifat sangat dinamis. Hal ini terjadi karena Random Forest menghasilkan prediksi akhir melalui penggabungan nilai rata-rata dari banyak pohon keputusan, sehingga hasil yang diperoleh cenderung lebih stabil, halus, dan tidak terlalu ekstrem.

Analisis Kontribusi Variabel Iklim Menggunakan SHAP

Dalam penelitian ini, nilai SHAP dihitung pada model yang menggunakan seluruh fitur input, tetapi interpretasi difokuskan pada kontribusi variabel iklim. Hasil kontribusi variabel iklim berdasarkan SHAP disajikan pada Tabel 3.

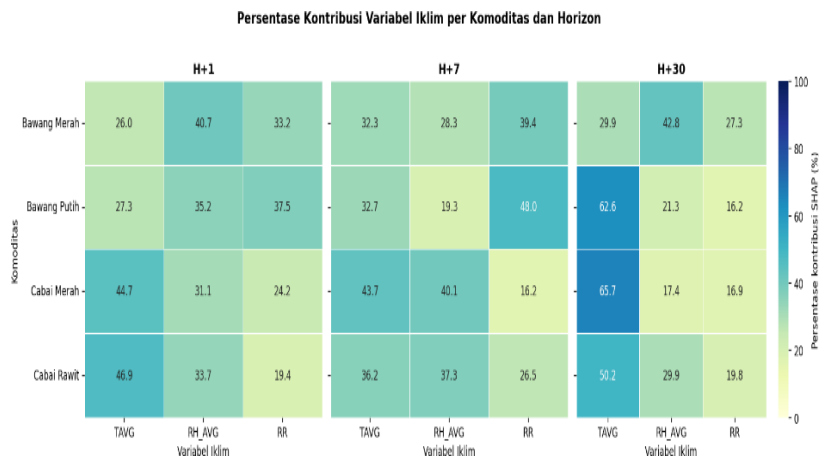
Tabel 3. Kontribusi Variabel Iklim Berdasarkan SHAP

Komoditas	Horizon	Variabel Iklim Dominan	Mean abs <i>SHAP</i>	Persentase Kontribusi
Bawang Merah	H+1	RH_AVG	534,83	40,73%
Bawang Merah	H+7	RR	772,61	39,41%
Bawang Merah	H+30	RH_AVG	2.166,83	42,78%
Bawang Putih	H+1	RR	320,97	37,50%
Bawang Putih	H+7	RR	640,00	48,00%
Bawang Putih	H+30	TAVG	1.640,17	62,57%
Cabai Merah	H+1	TAVG	396,71	44,71%
Cabai Merah	H+7	TAVG	1.030,77	43,70%
Cabai Merah	H+30	TAVG	4.202,57	65,71%
Cabai Rawit	H+1	TAVG	481,78	46,89%
Cabai Rawit	H+7	RH_AVG	753,35	37,33%
Cabai Rawit	H+30	TAVG	2.762,19	50,24%

Berdasarkan Tabel 3, variabel iklim dominan berbeda pada setiap komoditas dan horizon prediksi. Pada komoditas cabai merah, TAVG menjadi variabel iklim dominan pada seluruh horizon prediksi, yaitu H+1, H+7, dan H+30. Pada cabai rawit, TAVG menjadi variabel dominan pada horizon H+1 dan H+30, sedangkan RH_AVG dominan pada horizon H+7. Pola tersebut menunjukkan bahwa temperatur rata-rata memiliki kontribusi penting dalam prediksi harga kelompok cabai. Hasil ini sejalan dengan Ardian et al. (2024) yang menyatakan bahwa suhu menjadi salah satu variabel iklim

yang berpengaruh dalam prediksi harga cabai rawit berbasis iklim harian.

Pada kelompok bawang, kontribusi variabel iklim menunjukkan pola yang lebih bervariasi. Pada bawang merah, RH_AVG menjadi variabel dominan pada horizon H+1 dan H+30, sedangkan RR dominan pada horizon H+7. Pada bawang putih, RR menjadi variabel dominan pada horizon H+1 dan H+7, sedangkan TAVG menjadi variabel dominan pada horizon H+30. Perbedaan tersebut menunjukkan bahwa kontribusi variabel iklim terhadap prediksi harga tidak selalu sama pada seluruh komoditas. Setiap komoditas memiliki karakteristik biologis, pola pasokan, dan sensitivitas terhadap kondisi lingkungan yang berbeda sehingga respons model terhadap variabel iklim juga berbeda.



Gambar 3. Heatmap Kontribusi Variabel Iklim Berdasarkan SHAP

Gambar 3 memperlihatkan bahwa kontribusi variabel iklim terhadap model tidak bersifat seragam pada seluruh komoditas dan horizon prediksi. Pola tersebut menunjukkan bahwa setiap komoditas memiliki tingkat sensitivitas yang berbeda terhadap kondisi iklim, yang terlihat dari cara model memanfaatkan variabel input dalam proses prediksi pada kelompok cabai, variabel suhu cenderung menjadi informasi yang lebih dominan dalam pembentukan model, yang mengindikasikan bahwa perubahan temperatur memiliki keterkaitan yang kuat terhadap dinamika harga. Sementara itu, pada kelompok bawang, kontribusi variabel iklim lebih tersebar, membuktikan tidak ada satu variabel tunggal yang mendominasi secara tetap. Hal ini mencerminkan bahwa pembentukan harga pada komoditas bawang dipengaruhi oleh kombinasi beberapa faktor lingkungan. Temuan ini memperluas penelitian Ardian et al. (2024) melalui analisis empat komoditas hortikultura pada tiga horizon prediksi. Penelitian ini juga melengkapi Ibrahim (2025) dengan mengintegrasikan harga historis, variabel iklim, dan analisis SHAP, sehingga model tidak hanya menghasilkan prediksi harga, tetapi juga memberikan interpretasi mengenai kontribusi setiap variabel iklim terhadap hasil prediksi.

SIMPULAN

Berdasarkan hasil penelitian, *Random Forest Regression* dapat diterapkan untuk memprediksi harga komoditas hortikultura pangan strategis, yaitu bawang merah, bawang putih, cabai merah, dan cabai rawit di Provinsi Jawa Tengah dengan memanfaatkan harga historis dan variabel iklim pada horizon H+1, H+7, dan H+30. Hasil evaluasi menunjukkan bahwa kinerja terbaik terjadi pada horizon jangka pendek H+1 untuk seluruh komoditas, dengan performa terbaik pada bawang putih H+1 yang menghasilkan MAPE sebesar 1,69%, sedangkan galat tertinggi terjadi pada cabai merah H+30 dengan MAPE sebesar 26,72%. Analisis SHAP menunjukkan bahwa kontribusi variabel iklim berbeda pada setiap komoditas dan horizon prediksi, dengan TAVG dominan pada kelompok cabai, sedangkan pada kelompok bawang kontribusi variabel iklim lebih bervariasi antara RH_AVG, RR, dan TAVG. Dengan demikian, penggunaan Random Forest yang dikombinasikan dengan SHAP dapat menjadi pendekatan prediksi harga komoditas hortikultura yang tidak hanya menghasilkan nilai estimasi harga, tetapi juga mampu memberikan penjelasan mengenai kontribusi variabel iklim dalam model.

DAFTAR PUSTAKA

- Antad, S., Bag, V., Waghmode, O., Wattamwar, S., Wagh, A., & Zite, A. (2024). Kisan Dhan-Crop Price Prediction Using Random Forest. In *Communications on Applied Nonlinear Analysis* (Vol. 31).
- Ardian, M., Khomsah, S., Pandiya, R., Di, J., No, P., 128, K. P., Selatan, K., & Banyumas, J. (2024). *Perbandingan Model Regresi Untuk Memprediksi Harga Jual ...* 549. <https://hargajateng.org/tabel-harga->.
- Bifarin, O. O. (2023). Interpretable machine learning with treebased shapley additive explanations: Application to metabolomics datasets for binary classification. *PLoS ONE*, 18(5 May). <https://doi.org/10.1371/journal.pone.0284315>.
- Ibrahim, M. N. R. (2025). Price Forecasting of Shallots Using the Machine Learning Approach of Random Forest Regression Supporting Price Stabilization. *Jurnal Keteknik Pertanian*, 13(3), 449–461. <https://doi.org/10.19028/jtep.013.3.449-461>.
- Kaewchada, S., Ruang-On, S., Kuhapong, U., & Songsri-In, K. (2023). Random forest model for forecasting vegetable prices: a case study in Nakhon Si Thammarat Province, Thailand. *International Journal of Electrical and Computer Engineering*, 13(5), 5265–5272. <https://doi.org/10.11591/ijece.v13i5.pp5265-5272>
- Kumar, A., Kailasam, A. S., Rai, A., Khanna, M., Shukla, S., Das, S., & Chakraborti, A. (2025). *The Impact of Meteorological Factors on Crop Price Volatility in India: Case studies of Soybean and Brinjal*. <http://arxiv.org/abs/2503.11690>.
- Lestari, D. R., Bangun, E. A. S., Gaol, F. L., & Matsuo, T. (2025). Machine Learning-Based Forecasting of Agricultural Commodity Prices Using Ensemble Models. *Journal of Human, Earth, and Future*, 6(4), 887–899. <https://doi.org/10.28991/hef-2025-06-04-09>.
- Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *Stata Journal*, 20(1), 3–29. <https://doi.org/10.1177/1536867X20909688>.
- Ridwansyah Matondang, M., & Krisnamurthi, B. (2024). *Price Fluctuations and Volatility of National Strategic Food Commodities*. 8(1), 134. <http://ejournal2.undip.ac.id/index.php/agrisocionomics>.
- Talekar, B. (2020). A Detailed Review on Decision Tree and Random Forest. *Bioscience Biotechnology Research Communications*, 13(14), 245–248. <https://doi.org/10.21786/bbrc/13.14/57>.