Analisis Kinerja Model Hybrid Generalized Additive Model (GAM) dan Extreme Gradient Boosting (XGBoost) dalam Prediksi Harga Bawang Merah (Studi Kasus: Jawa Tengah)

Yoga Dwi Prasetyo[1], Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.[2], Trimono, S.Si., M.Si.[3]
[1, 3]*Afiliasi pertama (Sains Data, Universitas Pembangunan Nasional Veteran Jawa Timur)*
[1]22083010055@student.upnjatim.ac.id
[3]trimono.stat@upnjatim.ac.id
[2]*Afiliasi kedua (Sains Data, Universitas Pembangunan Nasional Veteran Jawa Timur)*
Corresponding author email: idhom@upnjatim.ac.id

ABSTRAK

Fluktuasi harga bawang merah di Jawa Tengah menimbulkan ketidakpastian pasar dan dapat merugikan petani maupun konsumen, sehingga diperlukan prediksi harga yang akurat sebagai dasar pengambilan keputusan. Penelitian ini menganalisis kinerja model hybrid Generalized Additive Model (GAM) dan Extreme Gradient Boosting (XGBoost) dalam memprediksi harga bawang merah berbasis data harian curah hujan periode 2020–2024. Hybrid dibangun dengan memanfaatkan output GAM sebagai fitur tambahan pada XGBoost, sehingga GAM menangkap pola non-linear jangka menengah akibat cuaca, sementara XGBoost mempelajari fluktuasi harga jangka pendek. Hasil eksplorasi menunjukkan bahwa curah hujan berpengaruh terhadap harga secara tertunda dengan korelasi tertinggi pada lag 4–6 hari. Evaluasi performa menunjukkan bahwa model hybrid menghasilkan tingkat kesalahan paling rendah dengan nilai MAPE sebesar 0,501%, lebih baik dibandingkan model tunggal GAM 0,797% maupun XGBoost 1,273%. Temuan ini menunjukkan bahwa integrasi model statistik dan machine learning efektif untuk meningkatkan akurasi prediksi harga komoditas hortikultura yang volatil, serta berpotensi mendukung sistem peringatan dini dan perencanaan strategi pasar berbasis data iklim.

Keywords: Prediksi harga, bawang merah, curah hujan, GAM, XGBoost, model hybrid.

I. PENDAHULUAN (10PT)

Fluktuasi harga bawang merah di Indonesia, khususnya di sentra produksi Jawa, berulang kali dilaporkan sebagai salah satu kontributor utama inflasi pangan dan sumber ketidakpastian bagi petani, pedagang, dan konsumen. Kajian kebijakan terbaru menunjukkan bahwa volatilitas harga bawang merah menuntut sistem peramalan yang andal agar pemerintah dan pelaku agribisnis dapat merancang strategi stabilisasi harga, pengelolaan stok, dan tata niaga yang lebih tepat sasaran [1]. Di berbagai daerah, kenaikan harga tajam sering muncul setelah periode panen yang terganggu, sementara penurunan harga yang dalam terjadi ketika produksi melimpah dan distribusi tidak tertata, sehingga kesejahteraan petani menjadi sangat rentan terhadap guncangan pasokan jangka pendek.

Sejumlah studi agroklimat dan perlindungan tanaman menegaskan bahwa salah satu sumber guncangan pasokan yang paling menonjol adalah variabilitas curah hujan dan anomali cuaca ekstrem. Analisis agroklimat di Tapanuli Selatan, misalnya, menemukan

bahwa perubahan pola hujan berpengaruh nyata terhadap produksi bawang merah dan penentuan kalender tanam, sehingga ketidaktepatan waktu tanam di bawah curah hujan yang tidak menentu dapat menurunkan hasil panen [2]. Studi lain mendokumentasikan bahwa anomali hujan tinggi pada 2021 di Lombok Barat meningkatkan hari hujan hingga 12 hari dan curah hujan 414% di atas normal, yang memicu serangan penyakit Fusarium dan menyebabkan kegagalan panen hampir total pada produksi benih bawang merah [3]. Penelitian di Jawa selama periode La Niña tiga tahun (2020–2022) juga menunjukkan bahwa iklim basah memperluas sebaran serangan Fusarium spp. pada bawang merah, dan secara tidak langsung berkontribusi terhadap fluktuasi harga melalui gangguan produksi [4].

Dari sisi pasar, analisis panel spasial harga kubis dan bawang merah di 34 provinsi Indonesia menunjukkan bahwa episode El Niño–La Niña, yang direpresentasikan oleh anomali curah hujan, berpengaruh signifikan terhadap pergerakan harga, terutama ketika kelebihan air tanah menyebabkan busuk umbi dan penurunan hasil [5]. Liputan kebijakan terkini menegaskan temuan ilmiah tersebut: Kementerian Pertanian dan Badan Riset dan Inovasi Nasional (BRIN) berulang kali mengingatkan bahwa curah hujan ekstrem dan perubahan pola musim meningkatkan risiko kegagalan panen bawang merah, sehingga mendorong upaya pengembangan varietas tahan iklim dan program peringatan dini untuk petani [6]. Dengan demikian, meskipun iklim tidak memengaruhi harga secara langsung, bukti empiris menunjukkan rantai hubungan curah hujan → produksi → pasokan → harga yang relevan bagi perencanaan pasar.

Di sisi metodologi peramalan, berbagai penelitian terkini telah mengevaluasi beragam pendekatan untuk memprediksi harga bawang merah di Indonesia. Pendekatan deret waktu klasik, seperti ARIMA dan SARIMA, masih banyak digunakan untuk menangkap pola musiman dan tren historis harga di berbagai pasar tradisional serta pada level produsen–konsumen [1]. Seiring berkembangnya komputasi, penelitian lain mengeksplorasi metode kecerdasan buatan seperti Fuzzy Inference System (FIS), jaringan saraf multilayer perceptron, dan Long Short Term Memory (LSTM) untuk memodelkan pola nonlinier pada harga bawang merah [7]. Hasil-hasil ini menunjukkan bahwa model nonlinier umumnya lebih adaptif terhadap fluktuasi tajam dibandingkan model linear, namun sebagian besar studi tersebut masih bertumpu pada satu jenis model dan mengandalkan data harga historis tanpa mengintegrasikan informasi iklim secara eksplisit.

Di saat yang sama, literatur iklim–pertanian menegaskan bahwa adaptasi budidaya bawang merah terhadap perubahan iklim mensyaratkan pemahaman kuantitatif mengenai respons tanaman terhadap curah hujan. Tinjauan terbaru mengenai budidaya bawang merah dalam konteks perubahan iklim menyoroti bahwa pola hujan yang bergeser dan kejadian hujan lebat memperbesar risiko genangan, busuk umbi, dan serangan penyakit, sehingga menekan produktivitas jika tidak diimbangi dengan teknologi budidaya adaptif [8]. Telaah lain memperjelas bahwa curah hujan dan kejadian La Niña yang memperparah kelembapan tanah menjadi pemicu penting fluktuasi hasil, yang kemudian tercermin dalam volatilitas harga di pasar [9]. Kombinasi bukti produksi dan harga ini memberikan landasan konseptual bahwa curah hujan layak dipertimbangkan sebagai variabel eksogen dalam model prediksi harga bawang merah, khususnya ketika data produksi tidak tersedia secara rinci.

Namun demikian, kajian yang secara khusus mengintegrasikan variabel iklim resolusi harian—seperti curah hujan—ke dalam model prediksi harga bawang merah masih relatif terbatas di Indonesia. Sebagian besar studi terdahulu berfokus pada model univariat dengan horizon data bulanan atau mingguan, sehingga belum sepenuhnya mengeksplorasi dinamika hubungan jangka pendek hingga menengah antara pola hujan harian dan penyesuaian harga di pasar daerah [1]. Selain itu, pemodelan yang mengedepankan akurasi sering kali mengorbankan transparansi, sehingga sulit bagi pembuat kebijakan untuk menafsirkan bagaimana perubahan curah hujan diterjemahkan

menjadi perubahan risiko harga.

Perkembangan terbaru dalam bidang machine learning interpretabel menawarkan peluang untuk menjembatani kebutuhan akurasi dan keterjelasan model. Generalized Additive Model (GAM) merepresentasikan kelas model aditif yang mengekspresikan keluaran sebagai jumlah fungsi halus (shape functions) satu-dimensi dari setiap prediktor, sehingga memungkinkan visualisasi langsung pengaruh marjinal tiap variabel terhadap respons [10]. Studi evaluasi mutakhir menunjukkan bahwa generasi baru GAM dan varian turunannya mampu menangkap pola nonlinier yang kompleks dengan performa prediksi yang kompetitif dibandingkan model black-box, sekaligus mempertahankan interpretabilitas tingkat fitur melalui kurva fungsi halus yang mudah dibaca [11].

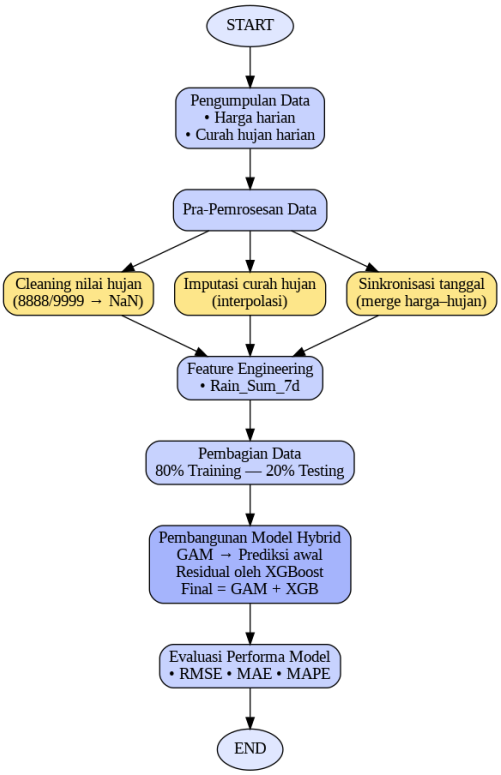
Di sisi lain, Extreme Gradient Boosting (XGBoost) telah menjadi salah satu algoritma pembelajaran ensemble berbasis pohon keputusan yang paling banyak digunakan untuk tugas regresi dan klasifikasi karena kemampuannya memodelkan hubungan nonlinier, menangani data dengan missing values, serta menggabungkan regularisasi dan komputasi paralel untuk meningkatkan akurasi dan efisiensi [12]. Sejumlah studi terkini menunjukkan bahwa XGBoost umumnya melampaui model baseline lain dalam berbagai konteks peramalan, mulai dari prediksi iklim hingga pemodelan deret waktu keuangan [13]. Selain itu, berbagai penelitian di bidang keuangan dan kesehatan telah mendemonstrasikan bahwa penggabungan XGBoost dengan model statistik atau metaheuristik dalam kerangka hibrida—misalnya GARCH–XGBoost atau LMD–ARIMA–XGBoost—dapat menghasilkan peningkatan akurasi yang konsisten dibandingkan model Tunggal [14]. Pendekatan serupa juga mulai diaplikasikan dalam studi lingkungan, di mana kombinasi GAM dan XGBoost dimanfaatkan untuk memetakan efek nonlinier sekaligus mengukur pentingnya variabel terhadap emisi dan indikator lingkungan lainnya [15].

Bertolak dari celah empiris dan metodologis tersebut, penelitian ini memfokuskan diri pada analisis kinerja model hybrid Generalized Additive Model–XGBoost untuk memprediksi harga harian bawang merah di Jawa Tengah dengan memasukkan curah hujan harian sebagai variabel eksogen utama. Secara khusus, pendekatan yang dikembangkan memanfaatkan kemampuan GAM untuk mempelajari dan memvisualisasikan hubungan halus antara agregat curah hujan (misalnya jendela rolling 1, 3, 7, dan 14 hari) dan harga, sekaligus menghasilkan output prediksi dan efek parsial yang kemudian digunakan sebagai fitur tambahan bagi model XGBoost. Dengan demikian, struktur stacked hybrid yang diusulkan memungkinkan XGBoost memanfaatkan informasi bentuk fungsi nonlinier yang telah "diringkas" oleh GAM, di samping mempelajari pola interaksi dan dinamika lag harga historis secara langsung.

Penelitian ini berkontribusi pada tiga aspek utama. Pertama, dari sisi substansi, studi ini memberikan bukti kuantitatif mengenai sejauh mana curah hujan berkontribusi terhadap variasi harga bawang merah harian di tingkat provinsi, bahkan ketika data produksi tidak tersedia secara eksplisit. Kedua, dari sisi metodologi, penelitian ini menghadirkan rancangan model stacked hybrid GAM–XGBoost yang menggabungkan transparansi fungsi halus dengan kekuatan prediktif gradient boosting, serta membandingkannya secara sistematis dengan model tunggal GAM dan XGBoost berdasarkan metrik RMSE, MAE, dan MAPE. Ketiga, dari sisi praktis, hasil penelitian diharapkan dapat menjadi dasar pengembangan sistem peringatan dini harga berbasis iklim yang lebih mudah diinterpretasikan oleh pemangku kepentingan, sehingga mendukung perencanaan tanam, manajemen risiko, dan kebijakan stabilisasi harga bawang merah di Jawa Tengah dan wilayah lain dengan karakteristik serupa.

II. METODOLOGI PENELITIAN (10 PT)

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksploratif untuk membangun sebuah sistem prediksi harga berbasis deret waktu dan informasi klimatologis. Tujuan utama penelitian adalah menghasilkan model prediksi harga bawang merah yang akurat sekaligus tetap interpretatif melalui integrasi pendekatan statistik dan machine learning. Adapun tahapan metodologi penelitian secara sistematis dijelaskan sebagai berikut:



Gambar 1. Tahapan Penelitian

II.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data harian dengan rentang periode 2020–2024. Terdapat dua dataset utama, yaitu:

1) Harga bawang merah harian, diunduh melalui sistem informasi harga pangan PIHPS Nasional. Data mencakup variabel tanggal dan harga komoditas bawang merah dalam satuan rupiah per kilogram.
2) Curah hujan harian, diunduh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Data memuat variabel tanggal dan curah hujan harian (RR) dalam satuan milimeter.

Kedua dataset diakses secara daring dari sumber resmi pemerintah.

II.2 Pra-Pemrosesan Data

Tahapan pra-pemrosesan dilakukan untuk memastikan kualitas data sebelum dilakukan pemodelan. Proses yang diterapkan meliputi:

1) Standarisasi format tanggal pada kedua dataset menggunakan tipe datetime.
2) Penyelarasan frekuensi observasi dengan menggabungkan dataset harga dan curah hujan berdasarkan tanggal (merge).
3) Pembersihan nilai anomali pada curah hujan, di mana kode 8888 (data tidak terukur) dan 9999 (tidak ada pengukuran) dikonversi menjadi NaN.
4) Imputasi nilai NaN menggunakan interpolasi deret waktu (time-series interpolation) untuk mempertahankan kelanjutan pola klimatologis.

Setelah tahap ini, diperoleh dataset harian yang konsisten dan siap digunakan untuk analisis deret waktu multivariat.

II.3 *Feature Engineering*

Feature Engineering dilakukan untuk merepresentasikan pengaruh akumulasi hujan terhadap dinamika harga komoditas. Fitur utama yang dibentuk adalah:

1) Rain_sum_1d, yaitu total curah hujan pada hari sebelumnya.
   Fitur ini digunakan untuk menangkap dampak perubahan cuaca yang terjadi sangat cepat, khususnya pada fase distribusi komoditas setelah panen.
2) Rain_sum_3d, yaitu akumulasi curah hujan selama 3 hari terakhir sebelum hari ke-t.
   Fitur ini menggambarkan tekanan klimatologis jangka pendek yang sering berkaitan dengan gangguan pasokan serta keterlambatan distribusi.
3) Rain_sum_7d, yaitu akumulasi curah hujan selama 7 hari terakhir sebelum hari ke-t.
   Fitur ini merepresentasikan tekanan klimatologis jangka menengah terkait proses pertumbuhan, panen, dan suplai yang masuk ke pasar.
4) Rain_sum_14d, yaitu akumulasi curah hujan selama 14 hari terakhir sebelum hari ke-t.
   Fitur ini digunakan untuk melihat kemungkinan efek residu curah hujan terhadap aktivitas produksi dan logistik.
5) Price_lag_1, yaitu harga bawang merah pada hari sebelumnya.
   Fitur ini menangkap karakteristik persistensi harga komoditas hortikultura, di mana harga hari ini sangat dipengaruhi oleh harga pada hari sebelumnya.

Setelah pembentukan seluruh fitur, dataset dibagi secara kronologis untuk menjaga integritas deret waktu, yaitu 80% untuk data pelatihan dan 20% untuk data pengujian, sehingga tidak terjadi kebocoran informasi dari masa depan ke model.

II.4 Pembangunan Model Hybrid GAM–XGBoost

Model hybrid dikembangkan dengan pendekatan stacked residual learning melalui dua tahapan utama:

1) Pemodelan awal menggunakan GAM, yang mempelajari pola hubungan non-linear antara curah hujan dan harga menggunakan spline penalized. Keluaran GAM berupa prediksi awal harga (GAM_output) serta nilai residual, yaitu selisih antara nilai aktual dan hasil prediksi GAM.
2) Pemodelan lanjutan menggunakan XGBoost, yang dilatih untuk memprediksi residual berdasarkan fitur awal (Rain_sum_7d) serta GAM_output. XGBoost berfungsi menangkap pola volatil dan interaksi non-linear yang tidak terwakili oleh GAM.

Hasil akhir model hybrid diperoleh dengan menjumlahkan GAM_output dan prediksi residual XGBoost. Integrasi ini memaksimalkan keunggulan kedua pendekatan: kemampuan GAM dalam memberikan interpretasi melalui fungsi halus (smooth functions) dan kemampuan XGBoost dalam mempelajari pola harga jangka pendek yang fluktuatif.

II.5 Evaluasi Performa Model

Kinerja prediksi diukur dengan tiga metrik evaluasi yang umum digunakan pada konteks prediksi harga komoditas:

1) Root Mean Square Error (RMSE) – mengukur tingkat kesalahan prediksi yang sensitif terhadap peningkatan harga ekstrem.
2) Root Mean Square Error (RMSE) – mengukur tingkat kesalahan prediksi yang sensitif terhadap peningkatan harga ekstrem.
3) Mean Absolute Percentage Error (MAPE) – mengukur akurasi relatif prediksi dalam konteks nilai harga aktual konsumen.

Evaluasi dilakukan dengan membandingkan tiga model: model hybrid GAM–XGBoost, model GAM tunggal, dan model XGBoost tunggal. Perbandingan performa ditampilkan dalam bentuk tabel serta grafik tren prediksi terhadap harga aktual pada Bab Hasil dan Analisis.

III. HASIL DAN ANALISIS

Bagian ini menyajikan hasil penelitian secara komprehensif mulai dari penggabungan data dan proses pra-pemrosesan, analisis eksploratif data, interpretabilitas model, hingga evaluasi performa prediktif model-model yang diuji. Seluruh analisis dilakukan menggunakan data harian harga bawang merah dan curah hujan periode tahun 2020–2024 di Jawa Tengah.

III.1 Hasil Pengumpulan Data dan Pra-pemrosesan

Dataset harga bawang merah diperoleh dari PIHPS dan dataset curah hujan diperoleh dari BMKG. Kedua dataset kemudian digabungkan berdasarkan tanggal sehingga diperoleh satu dataset terpadu berfrekuensi harian. Sebelum digunakan dalam pemodelan, dilakukan serangkaian pra-pemrosesan meliputi: penyamaan format tanggal, konversi satuan numerik, penanganan nilai hilang, eliminasi kode 8888/9999 pada curah hujan, dan imputasi interpolatif sesuai keterhubungan waktu.

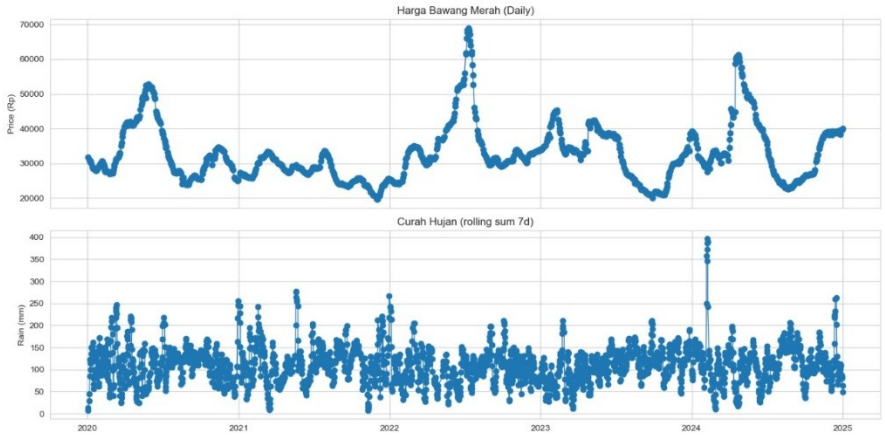
Tabel 1. Rekap hasil pra-pemrosesan dataset harga bawang merah dan curah hujan

date	price	price_lag_1	rain_sum_7d	rain_sum_7d_lag_1
2020-01-03	31700.0	32650.0	7.6	2.6
2020-01-03	31700.0	31700.0	12.6	7.6
2020-01-06	31000.0	31700.0	29.0	12.6
2020-01-06	31000.0	31000.0	45.4	29.0
2020-01-07	30500.0	31000.0	84.6	45.4
2020-01-07	30500.0	30500.0	122.5	84.6
2020-01-08	30600.0	30500.0	121.7	122.5
2020-01-08	30600.0	30600.0	117.2	121.7
2020-01-09	30500.0	30600.0	114.2	117.2
2020-01-09	30500.0	30500.0	99.8	114.2

Berdasarkan Tabel 1, seluruh observasi telah valid tanpa missing value setelah dilakukan imputasi kurasi. Selain itu, variabel turunan berupa rain_sum_1d, rain_sum_3d, rain_sum_7d, rain_sum_14d dan price_lag_1 ditambahkan untuk menangkap dinamika pergantian harga jangka pendek akibat shock pasokan. Penambahan fitur ini bertujuan memberikan konteks temporal bagi model prediksi.

III.2 Eksplorasi Data (EDA)

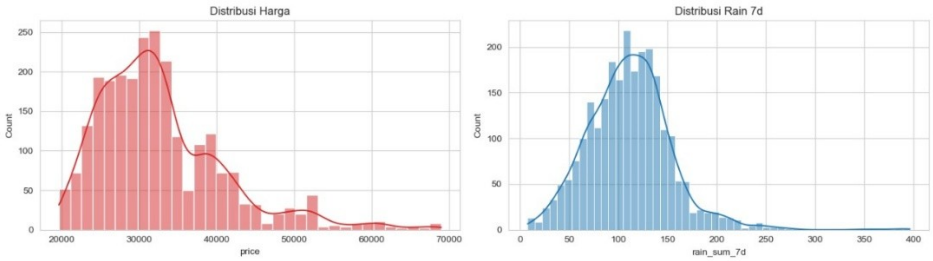
Langkah pertama dalam analisis adalah mengamati pola umum harga bawang merah dan curah hujan secara deret waktu.



Gambar 2. Tren harga bawang merah harian dan total curah hujan 7-hari (2020–2024)

Pada Gambar 2 terlihat bahwa harga bawang merah mengalami volatilitas tinggi setiap tahunnya, terutama pada Q2–Q3 2022 dan Q2 2023. Lonjakan harga tersebut mengindikasikan adanya shock pasokan yang kemungkinan disebabkan oleh musim penghujan dengan intensitas tinggi beberapa minggu sebelumnya. Di sisi lain, curah hujan menunjukkan pola musiman yang relatif konsisten: puncak pada awal tahun dan titik minimum pada pertengahan tahun. Perbedaan fase antara kurva hujan dan harga memperlihatkan bahwa hubungan keduanya tidak bersifat instan, tetapi memerlukan rentang waktu tertentu agar tercermin dalam harga pasar.

Distribusi harga dan curah hujan kemudian dianalisis untuk mengidentifikasi karakteristik statistik dasarnya.



Gambar 3. Distribusi harga bawang merah dan distribusi total curah hujan 7-hari

Pada Gambar 3 tampak bahwa distribusi harga bersifat right-skewed, dengan mayoritas harga berada pada rentang Rp 32.000 – Rp 38.000, namun terdapat ekor panjang dengan harga ekstrem melewati Rp 60.000. Distribusi ini mendukung temuan volatilitas harga komoditas hortikultura. Sementara itu, curah hujan cenderung berkumpul pada rentang intensitas rendah hingga menengah, namun tetap menunjukkan rentang ekstrem yang selaras dengan musim hujan tahunan. Pola statistik ini memberikan indikasi bahwa perubahan yang tajam pada harga dapat dipicu oleh kondisi hujan ekstrem, meskipun tidak setiap intensitas hujan tinggi pasti diikuti lonjakan harga.

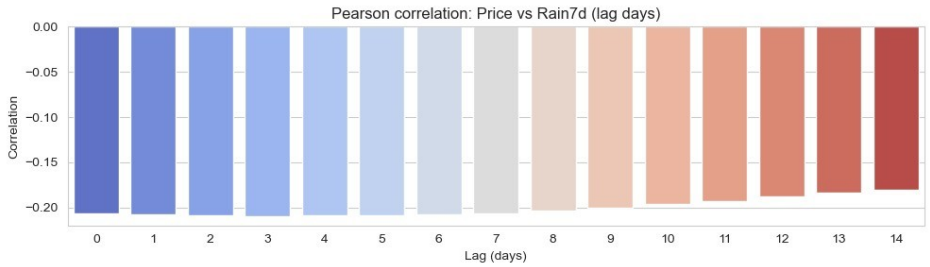
III.3 Analisis Hubungan Linier dan Non-Linier

Scatter plot dengan kurva LOESS digunakan untuk mengidentifikasi hubungan non-linier antara curah hujan dan harga.

Gambar 4. Scatter plot curah hujan 7-hari terhadap harga bawang merah dilengkapi kurva LOESS

Pada Gambar 4 terlihat bahwa harga bawang merah meningkat pada tren curah hujan rendah hingga sedang, namun setelah mencapai kisaran ±80 mm per minggu, kurva LOESS mulai mendatar. Hal ini menunjukkan bahwa curah hujan tidak terus meningkatkan harga setelah mencapai titik tertentu. Secara ekonomi, kondisi ini dapat dijelaskan karena intensitas hujan sangat tinggi dapat membuat produktivitas turun, tetapi pada titik tertentu pasokan yang berasal dari stok pasar dan inter-regency supply mampu menahan harga agar tidak terus naik.

Lebih jauh lagi, hubungan curah hujan dan harga dianalisis berdasarkan variasi waktu lag.

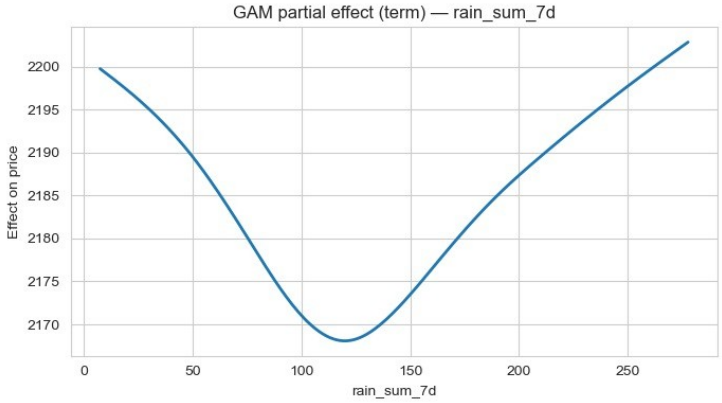


Gambar 5. Korelasi harga terhadap curah hujan dengan variasi lag 0–14 hari

Pada Gambar 5 korelasi tertinggi terjadi pada lag 4–6 hari, bukan pada lag 0. Temuan ini memperkuat kesimpulan bahwa curah hujan memengaruhi harga secara tertunda melalui mekanisme rantai pasok, bukan secara langsung. Dengan demikian, model prediksi harus mampu menangkap efek temporal jangka pendek sekaligus non-linier untuk meningkatkan akurasi prediksi.

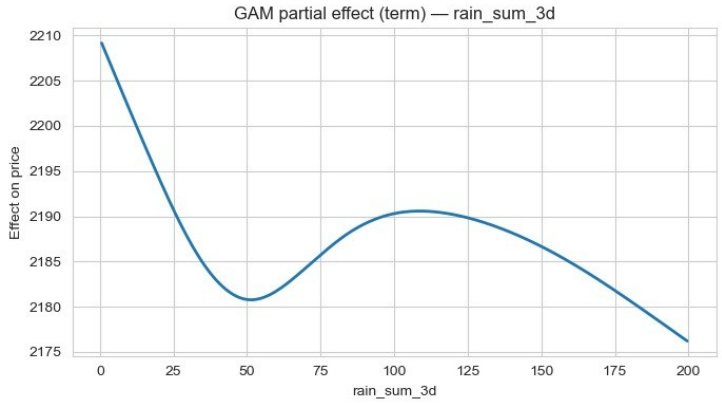
III.4 Analisis Interpretabilitas GAM

Model GAM digunakan untuk mengamati pola hubungan smooth antara variabel prediktor dan harga bawang merah. Tidak seluruh term bersifat signifikan sehingga hanya term signifikan yang ditampilkan secara visual.



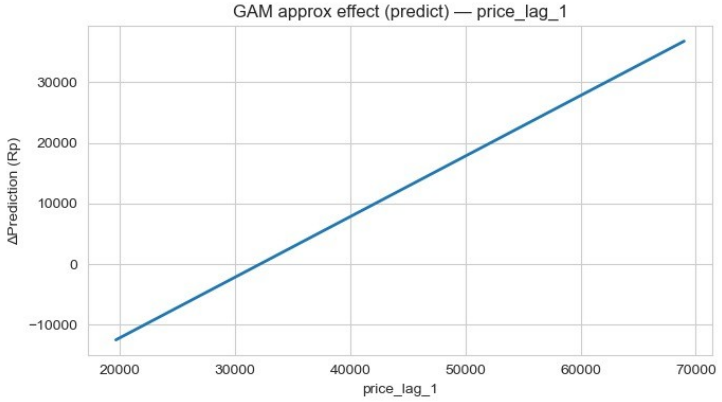
Gambar 6. Partial effect GAM untuk variabel rain_sum_7d

Pada Gambar 6 tampak bahwa harga bawang merah meningkat seiring kenaikan curah hujan hingga ±80 mm per minggu, setelah itu kurva menurun perlahan. Pola ini menunjukkan non-linearitas dan efek diminishing return, konsisten dengan perilaku pasar komoditas: ketika curah hujan menengah menyebabkan pasokan turun, harga naik, namun pada curah hujan ekstrem pelaku pasar mulai menahan permintaan sehingga kenaikan harga tidak berlanjut.



Gambar 7. Partial effect GAM untuk variabel rain_sum_3d

Gambar 7 menggambarkan bahwa curah hujan 3-hari memberikan efek yang lebih tajam dibanding rain_sum_7d. Hal ini menegaskan bahwa dinamika perubahan harga memiliki sensitivitas tinggi pada kondisional cuaca jangka pendek, dan fitur rain_sum_3d merupakan prediktor berpengaruh bagi model.



Gambar 8. Partial effect GAM untuk price_lag_1

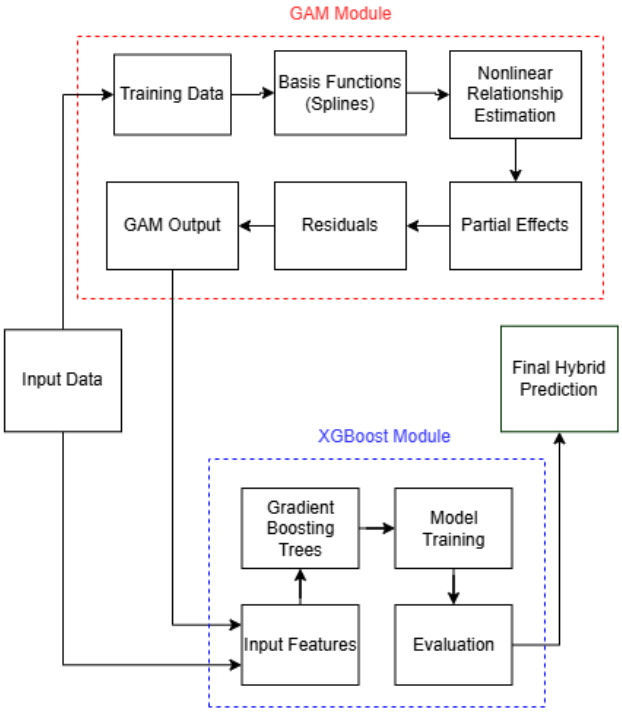
Gambar 8 memperlihatkan bahwa lag harga satu hari merupakan prediktor terkuat. Pola smooth menunjukkan garis searah (monoton rising), yang menegaskan bahwa harga bawang merah sangat bergantung pada harga periode sebelumnya. Temuan ini konsisten dengan teori bahwa volatilitas harga komoditas bersifat persistensial.

Term lain seperti rain_sum_1d dan rain_sum_14d menunjukkan pola rata (flat) dan tidak signifikan sehingga tidak ditampilkan karena tidak memberikan peningkatan interpretabilitas.

III.5 Performa Model Hybrid GAM–XGBoost

Model hybrid dibangun dengan menjadikan output prediksi GAM sebagai fitur tambahan untuk XGBoost, sehingga model dapat mempelajari dua komponen utama secara bersamaan: tren non-linear jangka menengah dari GAM dan lonjakan harga tajam yang dipelajari XGBoost.

Arsitektur Model Hybrid GAM–XGBoost ditampilkan pada Gambar 9, yang memvisualisasikan interaksi dua modul utama, yaitu GAM dan XGBoost. Modul GAM menerima data pelatihan untuk mempelajari hubungan non-linear antara fitur klimatologis dan harga, kemudian menghasilkan dua keluaran, yaitu GAM Output dan Residual. Residual tersebut selanjutnya menjadi salah satu fitur masukan bagi XGBoost untuk mempelajari pola-pola fluktuasi harga yang tidak dapat dijelaskan oleh GAM.



Gambar 9. Arsitektur Model Hybrid GAM–XGBoost

Penjelasan alur kerja ditunjukkan melalui dua segmentasi blok:

a) GAM Module (blok merah) melakukan basis transformation menggunakan spline, mempelajari hubungan non-linear antara harga dan variabel bebas, dan mengekstraksi partial effects dari masing-masing fitur iklim.

b) XGBoost Module (blok biru) menerima input berupa GAM Output, residual, serta variabel hasil feature engineering untuk meningkatkan akurasi prediksi melalui gradient boosting trees.

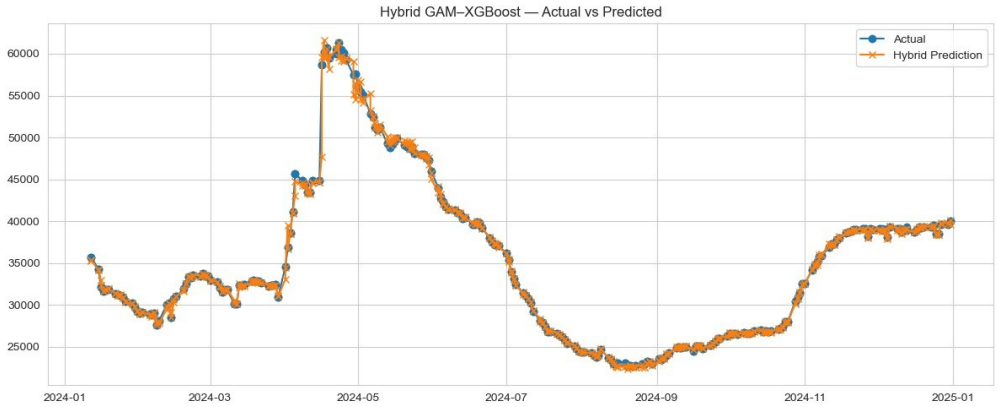
Integrasi keduanya memungkinkan model menangani dua permasalahan utama sekaligus:

1. Keterbatasan model statistik dalam mempelajari pola fluktuasi ekstrem, dan

2. Ketidakoptimalan model machine learning dalam menangkap karakteristik hubungan non-linear berbasis spline.

Dengan demikian, arsitektur ini dikembangkan untuk mengombinasikan kekuatan interpretabilitas GAM serta kekuatan prediktif XGBoost.

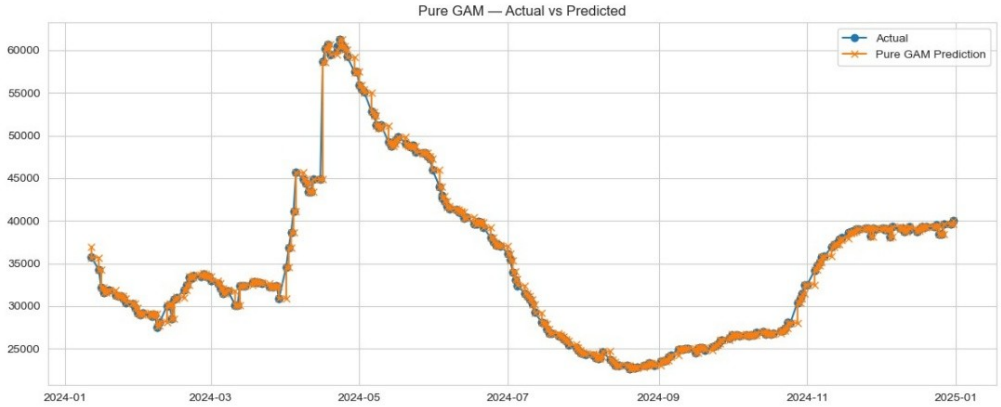
Selanjutnya, keluaran proses pemodelan GAM ditampilkan pada Gambar 10, yang menunjukkan grafik tren perbandingan harga aktual dan prediksi model Hybrid GAM–XGBoost.



Gambar 10. Perbandingan harga aktual dan prediksi model Hybrid GAM–XGBoost

Pada Gambar 10 tampak bahwa kurva prediksi model hybrid mengikuti pola harga aktual dengan ketepatan tinggi bahkan pada periode volatilitas ekstrem. Performa ini menunjukkan bahwa integrasi dua paradigma pemodelan memberikan keuntungan dalam kasus harga komoditas hortikultura yang dipengaruhi faktor eksogen secara tidak langsung.

III.6 Performa Model Pure GAM

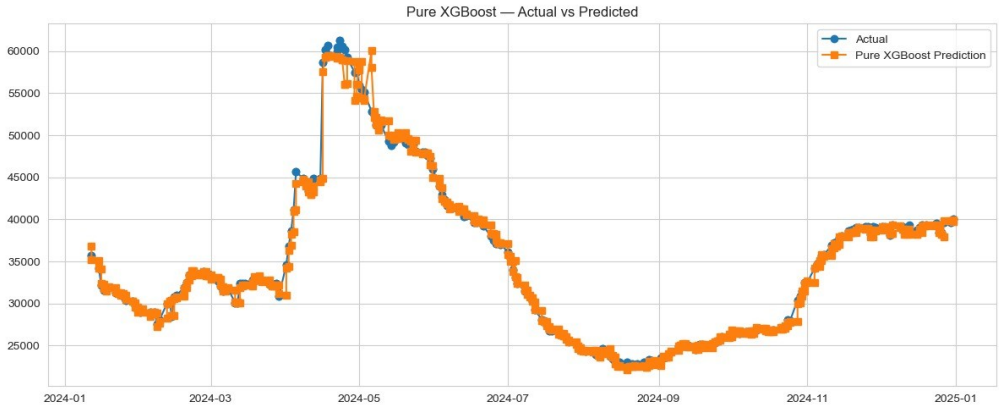


Gambar 11. Perbandingan harga aktual dan prediksi model Pure GAM

Pada Gambar 11 terlihat bahwa Pure GAM mampu memprediksi tren harga secara umum, namun tidak cukup responsif terhadap perubahan harga yang sangat cepat. Hal ini terjadi karena smoothing penalization menyebabkan model meredam fluktuasi tajam.

III.7 Performa Model Pure XGBoost

Pada Gambar 12 terlihat bahwa Pure XGBoost cukup baik mengikuti lonjakan harga, namun prediksinya kurang stabil pada periode harga stagnan atau moderat. Hal ini menunjukkan kecenderungan over-adaptation terhadap pola jangka pendek.



Gambar 12. Perbandingan harga aktual dan prediksi model Pure XGBoost

III.8 Perbandingan Kinerja Model

Semua hasil evaluasi prediksi dirangkum dalam tabel berikut.

Tabel 2. Perbandingan performa model Hybrid GAM–XGBoost, GAM, dan XGBoost

	MAE	RMSE	MAPE (%)
Hybrid (GAM + XGB)	204.169	621.924	0.501
Pure GAM	292.452	834.346	0.797
Pure XGBoost	490.270	1051.911	1.273

Berdasarkan Tabel 2, urutan performa terbaik adalah:

Hybrid GAM–XGBoost → Pure GAM → Pure XGBoost

Implikasi ekonomi dari temuan ini adalah bahwa harga bawang merah tidak dapat dijelaskan hanya oleh variabel cuaca maupun hanya oleh pola harga historis. Prediksi harga terbaik diperoleh ketika:

- pola smooth jangka menengah (efek curah hujan pada pasokan) ditangkap GAM
- pola lonjakan jangka pendek (sinyal pasar) ditangkap XGBoost.

Dengan demikian, kombinasi keduanya terbukti memberikan representasi struktural yang lebih lengkap dibanding pemakaian salah satu model saja.

IV. KESIMPULAN

Penelitian ini membuktikan bahwa curah hujan berpengaruh terhadap harga bawang merah secara non-linear dan tertunda, di mana intensitas hujan dalam beberapa hari sebelumnya berkontribusi terhadap perubahan harga melalui mekanisme pasokan. Hasil evaluasi model menunjukkan bahwa GAM unggul dalam menangkap pola smooth hubungan curah hujan terhadap harga, sedangkan XGBoost lebih responsif terhadap fluktuasi jangka pendek, namun masing-masing memiliki keterbatasan bila digunakan secara tunggal. Integrasi keduanya melalui model hybrid GAM–XGBoost menghasilkan performa prediksi terbaik, ditunjukkan oleh nilai MAE, RMSE, dan MAPE paling rendah dibandingkan model tunggal, sehingga model hybrid lebih representatif untuk memodelkan dinamika harga komoditas hortikultura yang volatil. Oleh karena itu, pemanfaatan model hybrid dapat menjadi pendekatan yang efektif untuk mendukung upaya pemantauan harga dan perencanaan strategi pasar pada komoditas bawang merah di Jawa Tengah.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Sains Data, UPN "Veteran" Jawa Timur atas dukungan yang diberikan dalam pelaksanaan penelitian ini. Ucapan terima kasih juga disampaikan kepada Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) serta Pusat Informasi Harga Pangan Strategis (PIHPS) atas penyediaan data yang menjadi dasar analisis penelitian ini. Penulis memberikan apresiasi kepada Tim DISCOVER selaku penyelenggara dan penyedia template prosiding ini. Semoga hasil penelitian ini dapat memberikan kontribusi dalam pengembangan ilmu pengetahuan, khususnya pada bidang analitik data harga komoditas pertanian.

REFERENSI

[1] E. Nurjati dan F. Susanti Wiryawan, "Strategi Meningkatkan Daya Saing Bawang Merah Melalui Prediksi Harga," *Jurnal Ilmu Pertanian Indonesia*, vol. 29, no. 3, hlm. 342–355, Mar 2024, doi: 10.18343/jipi.29.3.342.

[2] I. S. Harahap dan H. Soleh, "Analysis of Rainfall as an Agroclimatological Element on Production and Determining the Planting Season of Shallots (Allium Cepa L.) in South Tapanuli Regency," *Formosa Journal of Applied Sciences*, vol. 3, no. 2, hlm. 787–802, Mar 2024, doi: 10.55927/fjas.v3i2.8286.

[3] B. N. Hidayah, T. Sugianti, M. Mardiana, dan A. Pramudia, "The impact of weather anomalies on shallot seed production in West Lombok, Indonesia," dalam *E3S Web of Conferences*, EDP Sciences, Mar 2023. doi: 10.1051/e3sconf/202337303003.

[4] K. S. Dharma, "The Influence of Climate Factors on the Incidence Area of Fusarium spp. in Shallots on Java Island during the Triple-Dip La Niña (2020-2022)," *Jurnal Perlindungan Tanaman Indonesia*, vol. 28, no. 2, hlm. 104, Jul 2024, doi: 10.22146/jpti.85639.

[5] R. Fitriana, H. Siregar, dan L. Anggraeni, "The Impact of El Nino and La Nina Towards The Prices of Cabbage and Shallot in Indonesia," *Jurnal Manajemen dan Agribisnis*, Jul 2022, doi: 10.17358/jma.19.2.195.

[6] "BRIN develops extreme weather-resistant shallot variety - ANTARA News.". [Daring]. Tersedia pada: https://en.antaranews.com/news/285369/brin-develops-extreme-weather-resistant-shallot-variety

[7] N. Rofiq dan A. Salim, "RESOLUSI : Rekayasa Teknik Informatika dan Informasi Prediksi Harga Bawang Merah menggunakan Algoritma Fuzzy Inference System (FIS)", [Daring]. Tersedia pada: https://djournals.com/resolusi

[8] H. Rahmawan, N. I. D. Arista, S. Awanis, dan S. Z. A. Deli, "Resilient onions: Adapting onion cultivation to climate change," *Holistic: Journal of Tropical Agriculture Sciences*, vol. 3, no. 1, hlm. 59–78, Jul 2025, doi: 10.61511/hjtas.v3i1.2025.2135.

[9] K. S. Dharma, "The Influence of Climate Factors on the Incidence Area of Fusarium spp. in Shallots on Java Island during the Triple-Dip La Niña (2020-2022)," *Jurnal Perlindungan Tanaman Indonesia*, vol. 28, no. 2, hlm. 104, Jul 2024, doi: 10.22146/jpti.85639.

[10] "Additive model - Wikipedia.". [Daring]. Tersedia pada: https://en.wikipedia.org/wiki/Additive_model

[11] S. Kruschel, N. Hambauer, S. Weinzierl, S. Zilker, M. Kraus, dan P. Zschech, "Challenging the Performance-Interpretability Trade-Off: An Evaluation of Interpretable Machine Learning Models," *Business and Information Systems Engineering*, 2025, doi: 10.1007/s12599-024-00922-2.

[12] O. J. Abiodun dan A. I. Wreford, "Student's Performance Evaluation Using Ensemble Machine Learning Algorithms," *Engineering and Technology Journal*, 2024, doi: 10.47191/etj/v9i08.23.

[13] L. Alfaris *dkk.*, "Predicting Ocean Current Temperature Off the East Coast of America with XGBoost and Random Forest Algorithms Using Rstudio," *Ilmu Kelaut*, vol. 29, no. 2, hlm. 273–284, 2024, doi: 10.14710/ik.ijms.29.2.273-284.

[14] I. Maingo, T. Ravele, dan C. Sigauke, "A Fusion of Statistical and Machine Learning Methods: GARCH-XGBoost for Improved Volatility Modelling of the JSE Top40 Index," *International Journal of Financial Studies*, vol. 13, no. 3, Sep 2025, doi: 10.3390/ijfs13030155.

[15] S. Pan, W. T. de Vries, J. Liang, W. Chen, dan J. Guo, "Interpretable machine learning unveils greater environmental burden of vertical urban growth over horizontal sprawl in China," *Sustain Cities Soc*, vol. 131, hlm. 106787, Sep 2025, doi: 10.1016/J.SCS.2025.106787.