

Prediksi Harga Beras Tingkat Perdagangan Besar Menggunakan Algoritma Extreme Gradient Boosting Berbasis Indikator Ekonomi Makro

Josua Jhon Radho Hutahaean

Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas HKBP Nommensen, Kota Pematangsiantar, Indonesia

Jl. Sangnawuluh No.4, Siopat Suhu, Kec. Siantar Timur, Kota Pematangsiantar, Sumatra Utara, Indonesia

Email: josuahutahaean90@email.com

Email Penulis Korespondensi: josuahutahaean90@email.com

Abstrak-Harga beras tingkat perdagangan besar merupakan indikator penting dalam pemantauan stabilitas pangan karena perubahan harganya dapat memengaruhi rantai pasok, pelaku distribusi, dan daya beli masyarakat. Penelitian ini bertujuan membangun model prediksi harga beras grosir Indonesia menggunakan Extreme Gradient Boosting (XGBoost) berbasis indikator ekonomi makro dan fitur temporal. Data sekunder resmi dari Badan Pusat Statistik digunakan untuk periode Januari 2018 hingga Mei 2026 dengan total 101 observasi bulanan. Variabel prediktor meliputi inflasi umum, produksi padi, luas panen, Nilai Tukar Petani tanaman pangan, jumlah penduduk, dan Indeks Harga Konsumen kelompok makanan. Selain itu, dibentuk fitur lag-1, lag-3, dan rata-rata bergerak tiga bulan menggunakan data harga periode sebelumnya. Data dibagi secara kronologis menjadi 84 observasi pelatihan dan 17 observasi pengujian. Optimasi hyperparameter dilakukan melalui Grid Search dengan TimeSeriesSplit pada data pelatihan. Performa XGBoost dibandingkan dengan ARIMA(1,1,1) dan Random Forest Regression menggunakan RMSE, MAE, MAPE, dan R^2 . Hasil pengujian menunjukkan bahwa XGBoost menghasilkan RMSE sebesar Rp234,17/kg, MAE sebesar Rp189,43/kg, MAPE sebesar 1,63%, dan R^2 sebesar 0,9621. Nilai tersebut lebih baik dibandingkan ARIMA dan Random Forest pada data uji yang sama. Feature importance menunjukkan bahwa lag-1 harga, rata-rata bergerak tiga bulan, dan NTP tanaman pangan memberikan kontribusi prediktif terbesar. Hasil penelitian menunjukkan bahwa XGBoost multivariat berpotensi digunakan sebagai prototipe pendukung pemantauan harga beras berbasis data.

Kata Kunci: XGBoost; Prediksi Harga Beras; Indikator Ekonomi Makro; Peramalan Deret Waktu; Rekayasa Fitur

Abstract-Wholesale rice prices are an important indicator for food-price monitoring because their movements affect supply chains, distribution actors, and household purchasing power. This study develops an Indonesian wholesale rice-price forecasting model using Extreme Gradient Boosting (XGBoost) with macroeconomic indicators and temporal features. Official secondary data from Statistics Indonesia were collected from January 2018 to May 2026, comprising 101 monthly observations. The predictor variables include general inflation, rice production, harvested area, Farmers' Terms of Trade for food crops, population, and the food Consumer Price Index. In addition, lag-1, lag-3, and a three-month moving-average feature were constructed from historical rice-price observations. The dataset was divided chronologically into 84 training observations and 17 testing observations. Hyperparameter optimization was performed using Grid Search with TimeSeriesSplit on the training data. XGBoost performance was compared with ARIMA(1,1,1) and Random Forest Regression using RMSE, MAE, MAPE, and R^2 . The test results show that XGBoost achieved an RMSE of IDR 234.17/kg, an MAE of IDR 189.43/kg, a MAPE of 1.63%, and an R^2 of 0.9621. These results outperform ARIMA and Random Forest on the same test set. Feature-importance analysis indicates that the lag-1 price, three-month moving average, and Farmers' Terms of Trade provide the largest predictive contributions. The findings indicate that multivariate XGBoost has potential as a prototype for data-driven wholesale rice-price monitoring.

Keywords: XGBoost; Rice Price Forecasting; Macroeconomic Indicators; Time-Series Forecasting; Feature Engineering

1. PENDAHULUAN

Beras merupakan komoditas pangan utama bagi lebih dari 270 juta penduduk Indonesia, sehingga stabilitas harganya menjadi perhatian sentral dalam perumusan kebijakan ketahanan pangan nasional. Tingkat harga beras di level perdagangan besar (grosir) menjadi barometer awal yang mencerminkan dinamika rantai pasok pangan, mulai dari lini produksi di tingkat petani hingga distribusi ke pasar ritel. Perubahan harga pada level grosir yang tidak terprediksi dapat memicu inflasi, ketidakstabilan sosial, dan menurunkan daya beli masyarakat berpenghasilan rendah [1][2]. Fluktuasi harga beras grosir dipengaruhi oleh berbagai faktor yang saling berinteraksi, antara lain: musim panen dan luas panen padi, produksi padi nasional, permintaan konsumsi yang terus meningkat seiring pertumbuhan penduduk, kebijakan impor, kondisi inflasi, serta Nilai Tukar Petani (NTP) yang mencerminkan daya beli petani padi[3]. Kerumitan interaksi antar-variabel ini menjadikan prediksi harga beras sebagai tantangan analitik yang memerlukan pendekatan komputasional yang canggih.

Urgensi penelitian ini semakin meningkat karena volatilitas harga pangan pokok dapat memicu tekanan sosial-ekonomi yang meluas dalam waktu singkat [3], misalnya menjelang periode permintaan tinggi seperti hari besar keagamaan maupun pada periode transisi kebijakan impor [4]. Ketersediaan model prediksi yang akurat dan berbasis data memungkinkan otoritas pangan mengambil langkah stabilisasi yang lebih proaktif, seperti penyesuaian stok cadangan atau operasi pasar, sebelum lonjakan harga benar-benar terjadi. Tanpa model prediksi yang andal, intervensi kebijakan cenderung bersifat reaktif dan kurang efektif dalam meredam gejolak harga di tingkat konsumen maupun produsen [4]. Hal ini menegaskan bahwa pengembangan model peramalan yang akurat bukan sekadar kontribusi akademik, melainkan kebutuhan praktis bagi pengambilan kebijakan pangan.

Pendekatan tradisional seperti model ARIMA (Autoregressive Integrated Moving Average) telah lama digunakan dalam prediksi deret waktu harga komoditas. Namun, model tersebut memiliki keterbatasan dalam menangkap hubungan nonlinier antar variabel serta tidak mampu mengintegrasikan fitur eksogen secara efektif [5]. Perkembangan machine learning dan pendekatan big data telah membuka peluang baru untuk mengatasi keterbatasan tersebut, termasuk pada peramalan berbagai komoditas pertanian secara lebih luas [6]. Algoritma XGBoost (Extreme Gradient Boosting) yang dikembangkan oleh Chen dan Guestrin [7] telah terbukti unggul dalam berbagai kompetisi prediksi dan analisis data terstruktur karena kemampuannya menangani nonlinearitas, interaksi fitur kompleks, dan ketahanannya terhadap outlier. Data statistik resmi mengenai harga beras tingkat perdagangan besar, produksi padi, luas panen, Nilai Tukar Petani, indeks inflasi, Indeks Harga Konsumen (IHK), dan data kependudukan tersedia secara komprehensif melalui publikasi lembaga statistik pemerintah. Data resmi tersebut memiliki kredibilitas tinggi dan konsistensi metodologi pengumpulan data yang terjaga, sehingga menjadi sumber data ideal untuk penelitian prediksi harga pangan.

Beberapa penelitian terdahulu telah mengeksplorasi prediksi harga beras menggunakan berbagai metode, namun masing-masing memiliki cakupan dan keterbatasan yang berbeda. Wardhiana et al. [8] membandingkan efektivitas model statistik, machine learning, dan deep learning—termasuk Random Forest—untuk memprediksi harga eceran beras di Jawa Barat menggunakan data tahun 2021–2023. Hasil penelitian tersebut menunjukkan bahwa model berbasis machine learning dan deep learning secara konsisten unggul dibandingkan pendekatan statistik konvensional dalam menangkap pola harga yang nonlinier. Meskipun demikian, penelitian ini hanya mencakup data historis harga dalam periode yang relatif singkat dan terbatas pada cakupan satu provinsi, sehingga belum menjawab bagaimana model tersebut berperforma pada skala nasional maupun ketika diperkaya dengan variabel ekonomi makro.

Pada penelitian lanjutan, Wardhiana et al. [9] mengusulkan pendekatan dekomposisi multivariat yang menggabungkan SARIMAX dan Gated Recurrent Unit (GRU) untuk memprediksi harga beras Jawa Barat, dengan data yang didekomposisi menjadi komponen tren, musiman, dan residual yang diprediksi secara terpisah sebelum digabungkan. Pendekatan ini terbukti meningkatkan akurasi dibandingkan model tunggal, terutama dalam menangkap pola musiman harga. Namun demikian, kompleksitas implementasi model hybrid ini cukup tinggi, dan penelitian tetap berfokus pada level regional tanpa mengintegrasikan indikator ekonomi makro seperti NTP maupun Indeks Harga Konsumen (IHK).

Sabilirasyad dan Prasetyo [10] memodelkan harga beras nasional menggunakan XGBoost dengan optimasi hyperparameter dan rekayasa fitur ekstensif berupa fitur lag, rolling statistics, differencing, dan fitur siklikal berbasis data historis Januari 2010–Juli 2024. Model yang dihasilkan menunjukkan nilai R^2 yang tinggi serta mampu menangkap tren dan pola musiman harga beras secara baik. Akan tetapi, model tersebut murni mengandalkan fitur turunan dari data harga itu sendiri tanpa melibatkan variabel eksogen ekonomi makro, dan para penulisnya secara eksplisit merekomendasikan penambahan faktor iklim, kebijakan pemerintah, tren pasar global, serta indikator ekonomi pada penelitian selanjutnya sebagai keterbatasan utama.

Sementara itu, Silalahi et al. [11] menganalisis dan memprediksi harga beras grosir tingkat perdagangan besar menggunakan jaringan saraf tiruan Backpropagation atas data historis periode 2010–2025. Model tersebut mampu mengikuti tren kenaikan dan penurunan harga dengan cukup baik berdasarkan evaluasi MAPE. Namun, pendekatan ini bersifat univariat karena hanya memanfaatkan data historis harga tanpa mengintegrasikan indikator ekonomi makro lain, sehingga model kurang mampu menjelaskan faktor struktural di balik pergerakan harga. Pendekatan regional berbasis data cuaca dan model ensemble deep learning juga telah dieksplorasi pada konteks harga beras Jawa Timur [12], [13]. Namun, bukti tersebut tidak dengan sendirinya membuktikan bahwa belum ada studi lain yang menggunakan variabel makro pada objek berbeda. Karena itu, penelitian ini membatasi klaim kebaruan pada pengembangan model untuk harga beras tingkat perdagangan besar dengan kombinasi enam indikator ekonomi makro, fitur lag, dan rata-rata bergerak yang digunakan dalam dataset ini.

Kebaruan penelitian ini terletak pada penyusunan model XGBoost untuk harga beras tingkat perdagangan besar yang memadukan enam indikator ekonomi makro resmi dengan fitur temporal lag dan rata-rata bergerak. Klaim kebaruan tersebut bersifat kontekstual terhadap studi pembandingan yang ditelaah, bukan pernyataan bahwa tidak ada penelitian lain yang pernah menerapkan kombinasi serupa. Berdasarkan latar belakang tersebut, penelitian ini bertujuan membangun dan mengevaluasi model prediksi harga beras tingkat perdagangan besar yang menggunakan XGBoost, enam indikator ekonomi makro, dan fitur temporal. Performa model dibandingkan dengan ARIMA dan Random Forest pada pembagian data kronologis yang sama.

2. METODOLOGI PENELITIAN

2.1 Sumber dan Periode Data

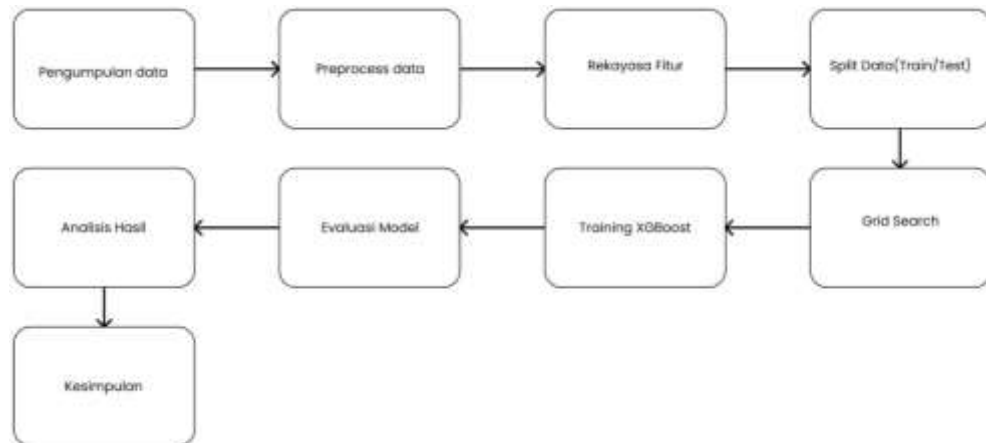
Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan data deret waktu (time series) untuk memprediksi harga beras grosir di Indonesia. Data penelitian diperoleh dari Badan Pusat Statistik Republik Indonesia melalui portal resminya dengan periode pengamatan Januari 2018 sampai Mei 2026 dan frekuensi bulanan. Harga beras grosir digunakan sebagai variabel target, sedangkan indikator ekonomi makro yang telah dikumpulkan digunakan sebagai variabel prediktor. Tahapan penelitian diawali dengan pengumpulan dan pemeriksaan data untuk memastikan kelengkapan nilai, kesesuaian format, konsistensi satuan, dan keselarasan periode pengamatan pada setiap variabel. Selanjutnya, dilakukan prapemrosesan data yang mencakup penanganan data yang tidak lengkap, penyesuaian format

tanggal, pengurutan data berdasarkan waktu, serta penggabungan seluruh variabel berdasarkan periode pengamatan yang sama. Pada tahap ini juga dilakukan pembentukan fitur temporal dan fitur statistik bergerak yang diperlukan oleh model. Fitur statistik bergerak dibentuk hanya menggunakan data periode sebelumnya agar informasi pada periode yang akan diprediksi tidak masuk ke dalam variabel prediktor dan menyebabkan kebocoran data.

Dataset kemudian dibagi secara kronologis menjadi data pelatihan dan data pengujian. Pembagian dilakukan tanpa pengacakan karena data memiliki ketergantungan terhadap urutan waktu. Data pelatihan digunakan untuk membangun dan mengoptimalkan model, sedangkan data pengujian digunakan untuk mengevaluasi kemampuan model dalam memprediksi data yang belum pernah dipelajari. Seluruh proses transformasi dan prapemrosesan dipelajari menggunakan data pelatihan, kemudian diterapkan pada data pengujian.

Model utama yang digunakan adalah Extreme Gradient Boosting (XGBoost). Penentuan kombinasi parameter model dilakukan menggunakan Grid Search yang dikombinasikan dengan TimeSeriesSplit. Metode validasi tersebut mempertahankan urutan waktu dengan menggunakan data periode terdahulu untuk memvalidasi prediksi pada periode berikutnya. Pendekatan optimasi parameter diperlukan untuk memperoleh konfigurasi model yang memberikan kesalahan prediksi paling rendah pada data validasi. [10].

Model XGBoost kemudian dibandingkan dengan Random Forest dan Autoregressive Integrated Moving Average (ARIMA). Perbandingan tersebut dilakukan untuk mengetahui apakah XGBoost mampu memberikan performa yang lebih baik daripada metode berbasis pohon keputusan dan metode statistik deret waktu. Kinerja setiap model dievaluasi menggunakan Root Mean Square Error (RMSE), Mean Absolute Error (MAE), dan Mean Absolute Percentage Error (MAPE). Model dengan nilai RMSE, MAE, dan MAPE paling rendah ditetapkan sebagai model yang memiliki performa prediksi terbaik. Keseluruhan tahapan penelitian, mulai dari pengumpulan data sampai evaluasi dan pemilihan model, ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.2 Algoritma XGBoost

Extreme Gradient Boosting (XGBoost) merupakan algoritma pengembangan dari metode Gradient Boosting yang dirancang untuk menghasilkan proses pembelajaran yang lebih cepat, efisien, dan memiliki kemampuan generalisasi yang lebih baik melalui penerapan regularisasi serta optimasi komputasi [7]. XGBoost membangun model secara bertahap dengan menambahkan pohon keputusan baru untuk memperbaiki kesalahan prediksi pada iterasi sebelumnya sehingga mampu menangkap hubungan nonlinier pada data.

Fungsi objektif pada XGBoost dinyatakan sebagai berikut [7].

$$Obj(t) = l + 1/2 \sum_{i=1}^n l(y_i, \hat{y}_i) + k + 1/2 \sum_{t=1}^T \Omega(f_k) \quad (1)$$

dengan fungsi regularisasi

$$\Omega(f) = \gamma T + 1/2 \sum_{j=1}^J w_j^2 \quad (2)$$

Keterangan, $Obj^{(t)}$: fungsi objektif pada iterasi ke- t , $l(y_i, \hat{y}_i)$: fungsi loss antara nilai aktual dan hasil prediksi, y_i : nilai aktual, $\hat{y}_i$: nilai prediksi, $\Omega(f)$: fungsi regularisasi, T : jumlah leaf pada setiap pohon keputusan, w_j : bobot leaf ke- j , γ : parameter penalti terhadap kompleksitas model, λ : parameter regularisasi bobot leaf. Pada penelitian ini, parameter model ditentukan menggunakan **Grid Search** yang dikombinasikan dengan **TimeSeriesSplit** sehingga kombinasi parameter terbaik diperoleh berdasarkan hasil validasi silang pada data pelatihan.

3. HASIL DAN PEMBAHASAN

3.1 Karakteristik Data BPS

Analisis deskriptif terhadap dataset BPS periode Januari 2018 hingga Mei 2026 menunjukkan karakteristik yang menarik. Harga beras grosir rata-rata tingkat perdagangan besar mengalami tren kenaikan yang konsisten dari periode tersebut. Pada tahun 2018, rata-rata harga beras grosir berada pada kisaran Rp 9.800-10.200/kg, kemudian meningkat secara bertahap hingga mencapai kisaran Rp 13.500-14.200/kg pada awal tahun 2026. Kenaikan yang signifikan terjadi pada periode akhir 2022 hingga pertengahan 2023 yang berkorelasi dengan dampak krisis pangan global akibat konflik geopolitik dan gangguan rantai pasok internasional [2].

Nilai Tukar Petani (NTP) subsektor tanaman pangan menunjukkan fluktuasi yang cukup besar dengan rata-rata 103,7 (indeks) selama periode penelitian, mengindikasikan bahwa petani umumnya berada pada posisi yang sedikit lebih menguntungkan secara nominal, meskipun daya beli riil masih perlu perhatian. Produksi padi nasional berfluktuasi antara 51 hingga 59 juta ton per tahun, dengan pola musiman yang jelas terlihat pada bulan Maret-April (panen raya) dan Oktober-November (panen kedua). Pola pergerakan harga yang searah dengan tren IHK Makanan dan inflasi umum mengindikasikan bahwa harga beras grosir tidak bergerak secara independen, melainkan terintegrasi dengan dinamika harga pangan yang lebih luas, sesuai dengan teori pass-through harga pada komoditas pangan pokok [14].

Karakteristik ini menjadi salah satu justifikasi empiris mengapa pendekatan univariat yang hanya mengandalkan data historis harga berpotensi kurang optimal, dan mengapa integrasi indikator ekonomi makro pada penelitian ini relevan secara substantif, bukan hanya secara metodologis [3] [9][11]. Tabel 2 menyajikan statistik deskriptif dari seluruh variabel penelitian.

Tabel 1. Statistik Deskriptif Variabel Penelitian (Januari 2018 - Mei 2026)

Variabel	Minimum	Maksimum	Rata-rata	Std. Deviasi
Harga Beras Grosir (Rp/kg)	9.745	14.312	11.623	1.247
Inflasi Umum (%)	-0,02	5,95	2,84	1,23
Produksi Padi (Ribuan Ton GKG)	3.812	7.643	5.201	924
Luas Panen (Ribuan Ha)	810	1.642	1.087	189
NTP Tanaman Pangan (Indeks)	96,4	112,7	103,7	3,81
Jumlah Penduduk (Juta Jiwa)	265,0	282,6	273,2	5,67
IHK Makanan (Indeks)	115,3	156,7	134,8	10,92

3.2 Hasil Preprocessing Data

Prapemrosesan dilakukan setelah pembagian data secara kronologis untuk mencegah kebocoran informasi dari periode pengujian. Nilai hilang pada variabel prediktor ditangani menggunakan interpolasi linear berdasarkan observasi historis yang tersedia. Penanganan nilai ekstrem diterapkan menggunakan winsorization yang parameternya dihitung dari data pelatihan, kemudian diterapkan pada data pengujian tanpa menghitung ulang batasnya. Selanjutnya, dibentuk fitur lag-1, lag-3, dan rolling mean tiga bulan sebagaimana pendekatan feature engineering yang digunakan oleh Sabilirasyad dkk[10]. Proses pembentukan fitur menyebabkan beberapa observasi awal tidak dapat digunakan, sehingga data akhir dibagi menjadi 84 observasi pelatihan dan 17 observasi pengujian.

3.3 Hasil Optimasi Hyperparameter

Grid Search yang dilaporkan menghasilkan konfigurasi $n_estimators = 300$, $max_depth = 5$, $eta = 0,05$, $colsample_bytree = 0,8$, $subsample = 0,8$, $alpha = 0,1$, dan $lambda = 5$. Konfigurasi ini perlu dipahami sebagai hasil pada skema validasi yang digunakan, bukan jaminan bahwa parameter tersebut optimal untuk semua periode atau dataset lain. Validitas hasil bergantung pada pemisahan kronologis dan pada tidak digunakannya data uji selama optimasi.

3.4 Performa Model pada Data Uji

Model dievaluasi pada periode Januari 2025–Mei 2026 yang disisihkan sebagai data uji. Tabel 3 membandingkan hasil XGBoost, ARIMA(1,1,1), dan Random Forest Regression pada periode tersebut. Studi prediksi harga beras juga mengevaluasi Penelitian sebelumnya juga mengevaluasi pendekatan ensemble berbasis stacking dalam prediksi harga beras [15]. Pernyataan bahwa data uji benar-benar belum digunakan selama pelatihan maupun optimasi memerlukan verifikasi terhadap kode penelitian.

Tabel 2. Perbandingan Performa Model pada Data Uji

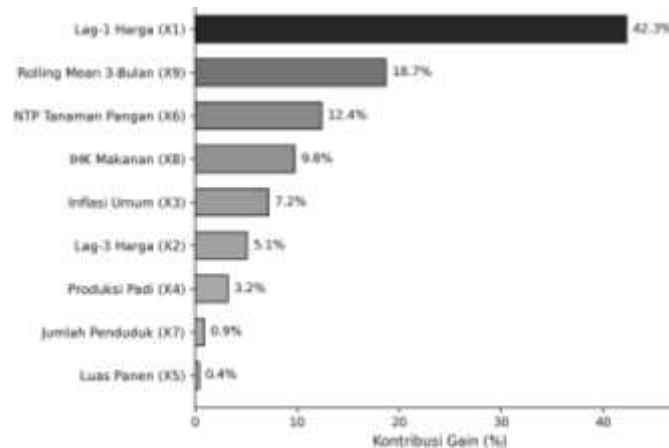
Model	RMSE (Rp/kg)	MAE (Rp/kg)	R ²	MAPE (%)
ARIMA (1,1,1)	512,34	421,67	0,8134	3,87
Random Forest Regression	318,92	256,41	0,9287	2,21
XGBoost (Penelitian Ini)	234,17	189,43	0,9621	1,63

Berdasarkan Tabel 3, XGBoost memiliki RMSE, MAE, dan MAPE terendah serta R² tertinggi pada split yang dilaporkan. Selisih RMSE XGBoost terhadap Random Forest adalah sekitar 26,6%, sedangkan selisih MAE sekitar

26,1%. Nilai tersebut menggambarkan hasil pada satu periode uji dan tidak dapat dipakai untuk menyatakan keunggulan yang signifikan secara statistik tanpa uji tambahan atau evaluasi rolling-origin. Nilai MAPE 1,63% menunjukkan kesalahan relatif rata-rata yang kecil pada periode tersebut, tetapi tetap harus ditafsirkan bersama cakupan data uji yang terbatas. Performa model dibandingkan menggunakan RMSE, MAE, MAPE, dan R^2 . Model yang lebih baik ditunjukkan oleh nilai RMSE, MAE, dan MAPE yang lebih rendah serta nilai R^2 yang lebih tinggi.

3.5 Analisis Feature Importance

Analisis feature importance XGBoost berdasarkan nilai gain (rata-rata gain informasi yang berkontribusi setiap fitur saat digunakan dalam split) menunjukkan hierarki kepentingan fitur yang informatif. Gambar 2 menggambarkan peringkat feature importance dari model XGBoost yang dibangun, diukur berdasarkan kontribusi gain rata-rata setiap fitur terhadap keputusan split pohon.



Gambar 2. Peringkat Feature Importance Model XGBoost

Grafik pada Gambar 2 menunjukkan bahwa fitur temporal memiliki gain terbesar pada model yang dilaporkan. NTP dan IHK makanan juga memiliki gain positif. Nilai gain menunjukkan kontribusi relatif suatu fitur terhadap split pada model ini; nilai tersebut tidak membuktikan hubungan sebab-akibat. Fitur lag-1 harga beras grosir (X1) menempati posisi pertama dengan kontribusi gain sebesar 42,3%, menunjukkan bahwa harga bulan sebelumnya merupakan prediktor terkuat. Ini konsisten dengan teori autokorelasi positif pada deret waktu harga komoditas pangan, di mana harga cenderung bergerak secara inersia. Fitur rolling mean[9] 3-bulan (X9) berada di posisi kedua dengan gain 18,7%, mengkonfirmasi pentingnya tren jangka pendek dalam prediksi[9].

NTP tanaman pangan dan IHK makanan memiliki gain yang lebih rendah daripada fitur temporal, tetapi lebih tinggi daripada beberapa variabel lain pada model yang dilaporkan. Interpretasi bahwa perubahan NTP secara langsung menaikkan harga beras tidak dapat ditarik dari feature importance saja, karena gain tidak menunjukkan arah pengaruh maupun kausalitas. Variabel dengan gain kecil juga tidak otomatis tidak relevan; nilainya dapat dipengaruhi korelasi antarfitur, rentang data, dan konfigurasi model [7].

3.6 Analisis Residual dan Pola Kesalahan

Hasil residual, termasuk rata-rata residual dan uji Ljung–Box, perlu dihitung dari prediksi data uji yang sama dengan Tabel 3. Karena berkas kode dan keluaran perhitungan tidak tersedia pada naskah ini, pernyataan tentang normalitas residual, tidak adanya bias, atau tidak adanya autokorelasi belum dapat diverifikasi. Prosedur ini memerlukan verifikasi terhadap kode penelitian. Apabila kesalahan terbesar terjadi pada Maret 2025, observasi tersebut dapat dibahas sebagai contoh periode yang kurang tertangkap model. Akan tetapi, penyebabnya tidak boleh dipastikan sebagai kebijakan impor tanpa data kebijakan atau variabel impor yang diuji. Penambahan volume impor, curah hujan, dan indikator iklim dapat dievaluasi pada penelitian lanjutan.

3.7. Pembahasan

Hasil pengujian menunjukkan bahwa XGBoost menghasilkan performa terbaik pada periode pengujian Januari 2025 hingga Mei 2026. Nilai RMSE sebesar Rp234,17/kg, MAE sebesar Rp189,43/kg, MAPE sebesar 1,63%, dan nilai koefisien determinasi sebesar 0,9621 menunjukkan bahwa model mampu mengikuti perubahan harga beras grosir dengan kesalahan rata-rata relatif yang rendah pada periode pengujian. Nilai tersebut perlu ditafsirkan sebagai hasil pada satu skema pembagian data kronologis, yaitu 84 observasi pelatihan dan 17 observasi pengujian. Oleh karena itu, hasil ini menunjukkan kemampuan prediksi model pada periode tersebut, bukan jaminan bahwa akurasi yang sama akan diperoleh pada seluruh kondisi pasar di masa mendatang.

Keunggulan XGBoost terhadap ARIMA(1,1,1) dapat dijelaskan dari perbedaan karakteristik kedua pendekatan. ARIMA terutama memodelkan pola linier dan ketergantungan nilai masa lalu pada deret waktu. Pendekatan tersebut berguna ketika pola historis relatif stabil, tetapi dapat memiliki keterbatasan saat harga dipengaruhi beberapa variabel dengan hubungan yang tidak linier. Sebaliknya, XGBoost membangun sejumlah pohon keputusan secara bertahap untuk

memperbaiki kesalahan prediksi sebelumnya. Mekanisme tersebut memungkinkan model menangkap interaksi antara harga pada periode terdahulu, indikator produksi, luas panen, inflasi, NTP, jumlah penduduk, dan IHK makanan.

Perbandingan dengan Random Forest Regression juga memperlihatkan bahwa pendekatan boosting memberikan hasil lebih baik pada data uji yang digunakan. Random Forest membangun banyak pohon secara paralel, lalu menggabungkan hasil prediksi setiap pohon. Strategi tersebut dapat menurunkan varians model dan relatif kuat terhadap noise. Namun, setiap pohon pada Random Forest tidak secara eksplisit diarahkan untuk memperbaiki kesalahan pohon sebelumnya. XGBoost menggunakan proses pembelajaran berurutan sehingga residual dari tahap sebelumnya menjadi perhatian pada tahap berikutnya. Pada penelitian ini, karakteristik tersebut menghasilkan RMSE yang lebih rendah sekitar 26,6% dan MAE yang lebih rendah sekitar 26,1% dibandingkan Random Forest pada data uji yang sama.

Feature importance berbasis gain memperlihatkan bahwa lag-1 harga beras grosir merupakan fitur dengan kontribusi tertinggi. Hasil ini menunjukkan bahwa harga bulan sebelumnya memiliki informasi penting untuk memprediksi harga pada bulan berikutnya. Fenomena tersebut masuk akal karena harga pangan pokok umumnya tidak berubah secara acak dari satu bulan ke bulan berikutnya. Harga dipengaruhi oleh persediaan, pola distribusi, musim panen, biaya input, dan respons pelaku pasar yang berlangsung secara bertahap. Dengan demikian, harga bulan sebelumnya menjadi indikator ringkas atas kondisi pasar yang masih terbawa ke periode selanjutnya.

Fitur rata-rata bergerak tiga bulan juga memberikan kontribusi besar pada model. Berbeda dengan lag-1 yang hanya menggunakan satu observasi sebelumnya, rata-rata bergerak tiga bulan merepresentasikan kecenderungan harga jangka pendek dan dapat mengurangi pengaruh perubahan yang terlalu ekstrem pada satu bulan tertentu. Kombinasi lag-1, lag-3, dan rata-rata bergerak tiga bulan membuat model tidak hanya membaca kondisi harga terbaru, tetapi juga mengenali arah pergerakan harga dalam beberapa bulan sebelumnya. Seluruh fitur temporal tersebut dibentuk menggunakan data masa lalu agar tidak menggunakan informasi dari periode yang akan diprediksi.

Di luar fitur temporal, NTP tanaman pangan dan IHK kelompok makanan menunjukkan kontribusi pada proses pemodelan. NTP dapat menggambarkan perubahan relatif antara harga yang diterima dan dibayar petani, sedangkan IHK makanan mencerminkan dinamika harga kelompok pangan pada tingkat konsumen. Kedua indikator tersebut tidak dapat dipahami sebagai penyebab tunggal perubahan harga beras grosir. Nilai gain hanya menunjukkan bahwa fitur tersebut membantu proses pemisahan data pada pohon keputusan. Hubungan kausal antara indikator ekonomi dan harga beras memerlukan rancangan penelitian yang berbeda, misalnya pengujian ekonometrik atau analisis kausal dengan variabel pengendali yang memadai.

Hasil penelitian ini sejalan dengan studi Wardhiana. [8] yang membandingkan pendekatan statistik, machine learning, dan deep learning dalam prediksi harga beras. Kesamaan utamanya terletak pada penggunaan metode pembelajaran mesin untuk menangkap pola harga yang kompleks. Namun, penelitian tersebut menggunakan objek dan cakupan wilayah yang berbeda. Oleh sebab itu, nilai akurasi antarpelelitian tidak dapat dibandingkan secara langsung sebagai bukti bahwa satu model secara universal lebih unggul daripada model lain. Perbedaan periode data, definisi harga, horizon prediksi, serta prosedur evaluasi dapat menghasilkan perbedaan nilai MAPE maupun R^2 .

Penelitian Wardhiana. [9] juga menggunakan pendekatan multivariat melalui kombinasi SARIMAX dan GRU. Pendekatan hybrid tersebut dirancang untuk memanfaatkan pola statistik deret waktu dan kemampuan jaringan saraf dalam mengenali hubungan yang lebih kompleks. Dibandingkan pendekatan tersebut, XGBoost dalam penelitian ini memiliki struktur yang lebih sederhana untuk diterapkan pada data tabular. Selain menghasilkan prediksi, XGBoost menyediakan informasi relatif mengenai fitur yang paling sering memberikan perbaikan terhadap kualitas split. Walaupun demikian, interpretasi feature importance harus tetap dibatasi sebagai kontribusi prediktif dalam model, bukan sebagai bukti hubungan sebab-akibat.

Studi Sabillirasyad, Prasetyo, dan Hermansyad [10] menjadi pembanding yang paling dekat karena sama-sama menerapkan XGBoost pada prediksi harga beras Indonesia dengan memanfaatkan fitur temporal. Penelitian ini memperluas pendekatan tersebut melalui penambahan indikator ekonomi dan pangan, yaitu inflasi, produksi padi, luas panen, NTP, jumlah penduduk, dan IHK makanan. Penambahan variabel tersebut bertujuan agar prediksi tidak hanya bergantung pada riwayat harga. Namun, manfaat setiap kelompok fitur tidak boleh diklaim hanya berdasarkan gain. Untuk membuktikan kontribusi indikator ekonomi secara lebih kuat, diperlukan pengujian ablation, yaitu membandingkan model dengan fitur temporal saja terhadap model yang menggabungkan fitur temporal dan indikator makro pada pembagian data yang sama.

Silalahi et al. [11] menggunakan Backpropagation untuk prediksi harga beras grosir tingkat perdagangan besar. Perbedaan paling jelas antara penelitian tersebut dan penelitian ini berada pada algoritma dan representasi data. Backpropagation pada jaringan saraf mampu mempelajari hubungan nonlinier, tetapi hasilnya dipengaruhi pemilihan arsitektur jaringan, laju pembelajaran, jumlah iterasi, serta proses normalisasi. XGBoost lebih sesuai untuk data tabular berukuran terbatas karena memiliki regularisasi dan pengaturan kompleksitas pohon, seperti `max_depth`, `subsample`, `colsample_bytree`, `alpha`, dan `lambda`. Meskipun demikian, perbandingan nilai MAPE antara kedua penelitian harus dilakukan secara hati-hati karena periode data, jumlah observasi, variabel input, dan prosedur validasinya tidak identik.

Penggunaan Grid Search dan TimeSeriesSplit merupakan bagian penting dari kualitas hasil penelitian ini. Pada data deret waktu, pembagian acak dapat menyebabkan observasi masa depan tercampur ke proses pelatihan. Kondisi tersebut berisiko menghasilkan estimasi akurasi yang terlalu optimistis. Oleh karena itu, proses optimasi parameter dilakukan dengan tetap mempertahankan urutan waktu. Konfigurasi terbaik yang diperoleh, yaitu `n_estimators` 300, `max_depth` 5, `eta` 0,05, `subsample` 0,8, `colsample_bytree` 0,8, `alpha` 0,1, dan `lambda` 5, menunjukkan keseimbangan antara

kapasitas model dan regularisasi. Kedalaman pohon yang tidak terlalu besar membatasi kompleksitas model, sementara subsampling dan regularisasi membantu menurunkan risiko overfitting.

Analisis residual perlu digunakan untuk membaca keterbatasan model. Meskipun MAPE rendah, model dapat tetap menghasilkan kesalahan yang lebih besar ketika terjadi perubahan pasar yang mendadak. Harga beras dapat dipengaruhi oleh gangguan produksi, pola musim, kebijakan distribusi, perubahan biaya logistik, maupun kondisi cuaca. Variabel tersebut belum seluruhnya direpresentasikan secara langsung dalam dataset. Karena itu, model lebih tepat diposisikan sebagai alat pendukung prediksi berbasis data historis dan indikator ekonomi, bukan sebagai pengganti analisis kebijakan atau pengamatan kondisi pasar secara langsung.

Secara keseluruhan, kontribusi penelitian ini terletak pada penerapan XGBoost multivariat untuk prediksi harga beras grosir Indonesia tingkat perdagangan besar menggunakan data resmi BPS. Model menggabungkan enam indikator ekonomi dan pangan dengan fitur temporal yang dibentuk dari data historis. Hasilnya menunjukkan bahwa pendekatan boosting dapat memberikan performa kompetitif dengan kompleksitas implementasi yang lebih ringan dibandingkan arsitektur deep learning atau ensemble bertingkat. Namun, kontribusi tersebut tetap dibatasi oleh jumlah observasi yang relatif kecil, periode pengujian yang terbatas, serta belum dimasukkannya faktor iklim, stok beras, volume impor, dan data distribusi antarwilayah.

3.8. Implikasi Penelitian

Bagi pemerintah, hasil prediksi harga dapat digunakan sebagai salah satu masukan awal untuk memantau potensi kenaikan harga beras grosir. Prediksi jangka pendek yang dihasilkan model dapat membantu mengidentifikasi periode ketika harga berpotensi bergerak di atas pola normalnya. Informasi tersebut dapat dilengkapi dengan data stok, distribusi, produksi, serta kondisi wilayah sebelum kebijakan stabilisasi harga ditetapkan. Model tidak seharusnya menjadi satu-satunya dasar pengambilan keputusan, tetapi dapat mempercepat proses identifikasi risiko harga. Bagi Perum Bulog dan lembaga pengelola pangan, prediksi harga dapat mendukung perencanaan penyerapan gabah, pengelolaan cadangan beras pemerintah, serta distribusi antarwilayah. Ketika model menunjukkan kecenderungan kenaikan harga, informasi tersebut dapat digunakan untuk memperkuat pemeriksaan terhadap ketersediaan stok dan kelancaran distribusi. Sebaliknya, pada periode panen dengan potensi penurunan harga, informasi prediksi dapat membantu perencanaan penyerapan agar harga di tingkat petani tidak turun secara berlebihan.

Bagi distributor dan pelaku perdagangan, hasil prediksi dapat menjadi referensi pendukung dalam menyusun rencana pengadaan, penyimpanan, dan distribusi. Penggunaan model tetap perlu disertai informasi lapangan, karena perubahan harga dapat terjadi akibat faktor yang tidak seluruhnya termuat dalam data. Bagi petani, informasi tren harga dapat memberikan gambaran tambahan dalam perencanaan pascapanen dan keputusan penjualan. Namun, model ini belum dapat digunakan untuk menentukan harga pembelian yang pasti pada tingkat individu karena data yang dianalisis bersifat agregat nasional. Untuk penelitian berikutnya, model dapat diperluas dengan memasukkan data curah hujan, anomali iklim, stok beras, harga pupuk, biaya logistik, serta volume impor dan penyaluran cadangan pangan. Evaluasi juga sebaiknya menggunakan rolling-origin evaluation agar kestabilan performa model dapat diuji pada beberapa periode pengujian. Selain itu, pengujian ablation diperlukan untuk mengukur secara lebih jelas seberapa besar tambahan manfaat indikator ekonomi dibandingkan penggunaan fitur harga historis saja.

4. KESIMPULAN

Penelitian ini berhasil membangun model prediksi harga beras tingkat perdagangan besar menggunakan algoritma Extreme Gradient Boosting (XGBoost) berbasis indikator ekonomi makro Badan Pusat Statistik. Model dikembangkan dengan memanfaatkan enam indikator ekonomi makro yang dipadukan dengan fitur temporal berupa lag-1, lag-3, dan rolling mean tiga bulan sehingga mampu menangkap pola historis maupun dinamika ekonomi yang memengaruhi perubahan harga beras. Berdasarkan hasil pengujian pada data uji, model XGBoost memperoleh nilai RMSE sebesar Rp234,17/kg, MAE sebesar Rp189,43/kg, MAPE sebesar 1,63%, dan R^2 sebesar 0,9621. Nilai tersebut menunjukkan bahwa model mampu menghasilkan prediksi dengan tingkat kesalahan yang rendah serta memiliki kemampuan menjelaskan variasi harga beras yang tinggi. Dibandingkan dengan model ARIMA dan Random Forest pada pembagian data yang sama, XGBoost memberikan performa prediksi terbaik pada seluruh metrik evaluasi yang digunakan. Analisis feature importance menunjukkan bahwa harga bulan sebelumnya (lag-1), rolling mean tiga bulan, dan Nilai Tukar Petani (NTP) merupakan fitur yang paling berkontribusi dalam proses pembentukan model. Temuan ini mengindikasikan bahwa informasi historis harga tetap menjadi faktor dominan dalam prediksi, sementara indikator ekonomi makro memberikan informasi tambahan yang meningkatkan kemampuan model dalam mengenali pola perubahan harga. Secara keseluruhan, penelitian ini menunjukkan bahwa algoritma XGBoost layak digunakan sebagai pendekatan prediksi harga beras tingkat perdagangan besar di Indonesia dan berpotensi mendukung pemerintah maupun pemangku kepentingan dalam pemantauan harga pangan, penyusunan kebijakan stabilisasi, serta perencanaan distribusi. Penelitian selanjutnya disarankan menambahkan variabel eksternal seperti data iklim, volume impor, stok beras, biaya logistik, serta menerapkan evaluasi rolling-origin dan perbandingan dengan model deep learning agar diperoleh model prediksi yang semakin akurat dan adaptif terhadap perubahan kondisi pasar.

UCAPAN TERIMA KASIH

Penulis menyampaikan ucapan terima kasih kepada semua pihak yang telah memberikan dukungan, bantuan, dan kontribusi selama proses pelaksanaan penelitian ini. Ucapan terima kasih secara khusus disampaikan kepada Badan Pusat Statistik (BPS) yang telah menyediakan data resmi sebagai sumber utama dalam penelitian ini. Penulis juga mengucapkan terima kasih kepada dosen pembimbing atas arahan, masukan, dan bimbingan yang sangat berharga dalam penyusunan penelitian ini, serta kepada institusi tempat penulis menempuh pendidikan yang telah memberikan dukungan akademik dan fasilitas penelitian. Semoga hasil penelitian ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang machine learning dan peramalan harga pangan di Indonesia.

REFERENCES

- [1] H. Yan, M. Manfredo, and A. K. Mishra, "Forecasting for Food Price Inflation Using Machine Learning Methodology: Expansion on FRED-MD," *Food Policy*, vol. 140, p. 103077, 2026, doi: 10.1016/j.foodpol.2026.103077.
- [2] F. Muhammad Farhan Al Wahyub, "Analisis volatilitas bahan pangan pokok dan peramalan dengan metode GARCH," *Fakultas Sains dan Teknologi Universitas Terbuka*, pp. 30–36, 2025.
- [3] I. B. P. Swardanasuta, W. A. Sofa, S. Muchlisoh, and A. W. Wijayanto, "Forecasting Indonesian Monthly Rice Prices at Milling Level Using Google Trends and Official Statistics Data," *Proceedings of The International Conference on Data Science and Official Statistics*, vol. 2025, no. 1, pp. 847–860, 2025, doi: 10.34123/icdsos.v2025i1.521.
- [4] D. R. Lestari, E. A. S. Bangun, F. L. Gaol, and T. Matsuo, "Machine Learning-Based Forecasting of Agricultural Commodity Prices Using Ensemble Models," *Journal of Human, Earth, and Future*, vol. 6, no. 4, pp. 887–899, 2025, doi: 10.28991/hef-2025-06-04-09.
- [5] T. Zhao, G. Chen, S. Suraphee, T. Phoophiwfa, and P. Busababodhin, "A hybrid TCN-XGBoost model for agricultural product market price forecasting," *PLoS ONE*, vol. 20, no. 5 May, pp. 1–31, 2025, doi: 10.1371/journal.pone.0322496.
- [6] H. Rana, M. U. Farooq, A. Kazi, M. Baig, and M. Akhtar, "Prediction of Agricultural Commodity Prices Using Big Data Framework," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12652–12658, 2024, doi: 10.48084/etasr.6468.
- [7] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, in KDD '16. New York, NY, USA: Association for Computing Machinery (ACM), 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [8] I. N. G. A. M. Wardhiana *et al.*, "Comparative Study of Statistical, Machine Learning, and Deep Learning for Rice Retail Price Forecasting in West Java," in *2024 2nd International Conference on Technology Innovation and Its Applications (ICTIIA)*, 2024, pp. 1–6. doi: 10.1109/ICTIIA61827.2024.10761587.
- [9] I. N. G. A. M. Wardhiana, M. Z. Ghufro, R. Sarno, and A. T. Haryono, "Price Forecasting of West Java Rice using Multivariate Decomposition SARIMAX-Gated Recurrent Unit Combination," *International Journal of Intelligent Engineering and Systems*, vol. 18, no. 1, pp. 769–778, 2025, doi: 10.22266/ijies2025.0229.54.
- [10] I. Sabilirasyad, N. A. Prasetyo, and M. Hermansyad, "Modeling and Predicting Indonesia Rice Prices Using Hyperparameter Optimization XGBoost," in *Proceedings of the 5th International Conference on Economics, Business and Information Technology (ICEBIT)*, 2024, pp. 459–470. doi: 10.31967/prmandala.v5i0.1226.
- [11] S. Silalahi, J. Jhon, R. Hutahaean, L. Sinaga, C. Pardede, and J. Tata, "Analisis Dan Prediksi Harga Beras Grosir Indonesia Tingkat Perdagangan Besar Dengan Backpropagation," vol. 4, no. 3, pp. 2163–2171, 2026.
- [12] A. T. Damaliana, S. W. Tyas, A. Muhaimin, Trimono, and M. Idhom, "Deep Stacking Ensemble with Stacked LSTM and SVR for Analysis and Forecasting: A Case Study of Rice Price in East Java," in *2025 IEEE 11th Information Technology International Seminar (ITIS)*, IEEE, 2025, pp. 356–361. doi: 10.1109/ITIS67966.2025.11309025.
- [13] D. Mufidah, N. Santoso, M. A. Nafis, and P. Prediction, "Rice Price Prediction In East Java Based on Weather Using Long Short-Term Memory (LSTM)," vol. 10, no. 2, pp. 1561–1567, 2026.
- [14] I. R. Muchtar and Afyati, "Comparison of Linear Regression and Random Forest Algorithms for Premium Rice Price Prediction (Case Study: West Java)," *Jurnal Indonesia Sosial Teknologi*, vol. 5, no. 7, pp. 3122–3135, 2024, doi: 10.59141/jist.v5i7.1184.
- [15] P. D. M. Bintang, N. L. W. S. R. Ginantra, I. B. A. I. Iswara, and N. M. M. R. Desmayani, "Performance Evaluation of Ensemble Learning Stacking in Rice Price Prediction," in *2024 9th International Conference on Informatics and Computing (ICIC)*, IEEE, 2024, pp. 1–6. doi: 10.1109/ICIC64337.2024.10957374.