

Implementation of Stacking Ensemble Learning on Decision Tree Regressor for Food Commodity Price Prediction in Indonesia

Nukman Solikhudin ^{1*}, Mega Novita ^{2*}, Ramadhan Renaldy ^{3*}

* Informatika, Universitas Persatuan Guru Republik Indonesia Semarang
nukmansn@gmail.com ¹, novita@upgris.ac.id ², ramadhanrenaldy@upgris.ac.id ³

Article Info

Article history:

Received 2026-05-25

Revised 2026-07-24

Accepted 2026-08-12

Keyword:

*Decision Tree Regressor,
Ensemble Learning,
Temporal Split,
Food Commodities,
Food Price Prediction,
Stacking*

ABSTRACT

Fluctuations in staple food prices across Indonesian regions exhibit complex, non-linear patterns vulnerable to market shocks. This study aims to construct an accurate, stable food price prediction model utilizing a Stacking Ensemble Learning approach. A raw dataset of 27,722 records from the National Food Agency was cleaned by removing invalid data and zero values, yielding 27,270 well-indexed observations. To address severe scale disparity between commodities and heteroscedasticity effects, a natural logarithm transformation was applied to the target variable. Time-series features, specifically Lag 1 and Moving Average 3, were locally constructed based on commodity-province groups to capture temporal dependencies. The proposed Stacking Ensemble model integrates four multi-architecture base learners Ridge Regression, AdaBoost, Gradient Boosting, and Extra Tree with a Decision Tree Regressor acting as the meta-learner. Model evaluation was conducted using a temporal split method with an 80:20 ratio to strictly prevent data leakage. Experimental results demonstrate that the proposed Stacking Ensemble model achieves superior performance on nominal test data compared to baseline models, securing an R^2 of 0.895, RMSE of 2,531, and MAE of 1,461. Furthermore, the model proved highly robust in balancing bias and variance, yielding the smallest R^2 Gap of 0.035. Model transparency analysis reveals a powerful temporal inertia, where historical features dominate the decision weight by up to 87.55%. However, per-commodity performance analysis highlights a performance limitation on subsidized commodities (Minyak Kita) due to data distortion caused by non-market Price Ceiling regulations. This study provides critical implications for food authorities to formulate data-driven, responsive market interventions.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Sandang, pangan, dan papan merupakan kebutuhan dasar manusia. Seiring dengan perkembangan zaman yang menyebabkan semakin modern dunia, kebutuhan manusia menjadi semakin beragam. Komoditas pangan merupakan salah satu kebutuhan dasar yang memiliki peranan penting dalam menunjang keberlangsungan hidup masyarakat. Di Indonesia, harga komoditas pangan cenderung tidak stabil dan kerap mengalami fluktuasi dari waktu ke waktu.[1]. Penentuan harga tersebut dipengaruhi oleh faktor pasokan dan permintaan. Pada periode tertentu, seperti saat perayaan hari besar, harga komoditas pangan berpotensi meningkat yang

dapat memicu munculnya terjadi inflasi harga komoditas pangan[2].

Pemerintah melakukan berbagai metode untuk antisipatif untuk menjaga kestabilan harga, salah satunya melalui kebijakan Harga Eceran Tertinggi (HET). Penentuan HET yang tepat memerlukan keahlian dalam memprediksi harga komoditas pangan sejak dini[3]. Seiring dengan perkembangan teknologi, pendekatan *Machine Learning* mulai banyak digunakan dalam prediksi harga karena kemampuannya dalam mengolah data. Teknologi ini mengharuskan sistem untuk mempelajari pola dari data historis sehingga dapat menghasilkan prediksi secara otomatis[4]. Salah satu metode dari Machine Learning yang

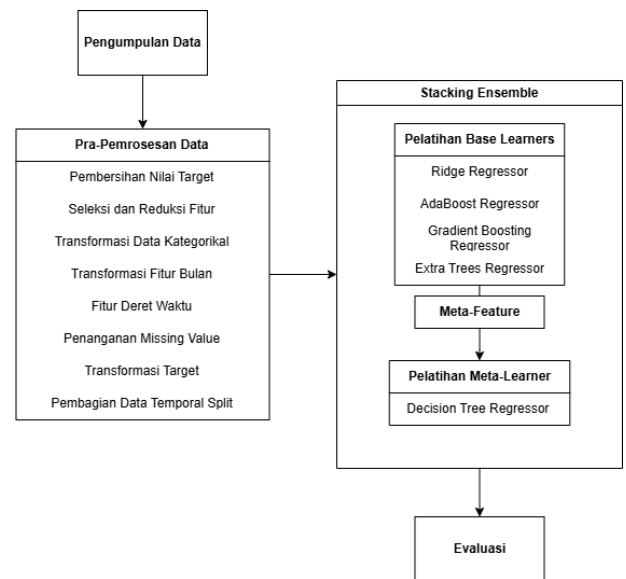
sedang berkembang adalah Ensemble Learning, yaitu metode yang menggabungkan beberapa model prediksi untuk meningkatkan kinerja, akurasi, dan stabilitas hasil prediksi [5]. Teknik stacking adalah salah satu Teknik dalam Ensemble Learning, menggabungkan beberapa model pembelajaran mesin tunggal sebagai base learners dan algoritma machine learning lainnya sebagai meta learner[6]. Beberapa penelitian menggunakan teknik stacking, Tirta et al. (2024) penelitian menunjukkan bahwa model GBM memberikan kinerja terbaik sebagai algoritma dasar dengan nilai MAPE 2.327 dan R^2 sebesar 0.973. Namun, ketika menggunakan teknik stacked ensemble yang menggabungkan beberapa model dasar, akurasi meningkat menjadi MAPE 2.299 dengan R^2 tetap 0.973[7]. Hibatulloh et al. (2025) Hasil studi komparatif menunjukkan bahwa Stacking Ensemble memberikan performa terbaik dengan R^2 0.9076, RMSE 0.311, dan MAE 0.216, mengungguli model tunggal seperti Random Forest dan XGBoost[8]. Dari penelitian terdahulu dapat meningkatkan akurasi dari prediksi.

Dengan demikian, penelitian ini bertujuan untuk mengimplementasikan metode Ensemble Learning dengan teknik stacking dalam memprediksi harga komoditas pangan di Indonesia. Model yang digunakan menggabungkan algoritma *Ridge Regression*, *AdaBoost Regressor*, *Gradient Boosting Regressor*, dan *Extra Trees Regressor* sebagai *base learners*, serta *Decision Tree Regressor* sebagai *meta-learner*. Kinerja model akan dievaluasi menggunakan RMSE dan R^2 . Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam mendukung kebijakan pen-gendalian harga pangan, perencanaan distribusi, serta pengendalian inflasi berbasis data prediktif.

II. METODE

Penelitian ini menggunakan metode kuantitatif dengan penerapan algoritma *machine learning* berbasis *ensemble learning* untuk memprediksi harga komoditas pangan di Indonesia. Pendekatan ini dipilih karena mampu mengolah data numerik dalam jumlah besar serta dapat meningkatkan akurasi prediksi melalui penggabungan beberapa model pembelajaran. Tahapan penelitian terdiri dari empat bagian utama yaitu pengumpulan data, pra-pemrosesan data, *Stacking Ensemble*, dan evaluasi model, sebagaimana ditunjukkan pada Gambar 1.

Data yang digunakan dalam penelitian ini diperoleh dari sumber terbuka Badan Pangan Nasional[9]. Dataset mencakup 21 komoditas pangan utama yang memiliki peranan penting dalam kebutuhan konsumsi masyarakat, di antaranya Beras Medium, Cabai Rawit Merah, Daging Ayam Ras, dan Minyak Goreng. Data harga komoditas dikumpulkan dari 39 provinsi di Indonesia, sehingga mampu merepresentasikan variasi kondisi harga antar wilayah secara nasional.



Gambar 1. Alur pelaksanaan Penelitian

Total 27.722 data harga komoditas pangan bulanan tahun 2021 sampai 2025 dengan beberapa variabel penting seperti nama kode provinsi, Nama provinsi, Komoditas, Tahun, Bulan dan harga. Tabel 1 berikut adalah sample dari dataset yang menunjukkan adanya variasi harga antarprovinsi. Perbedaan harga ini juga dapat dipengaruhi oleh ketersediaan pasokan lokal, tingkat konsumsi masyarakat, serta dinamika permintaan musiman yang berbeda pada setiap wilayah[10]. Dengan demikian, dataset tersebut memberikan gambaran yang kaya mengenai pola pergerakan harga komoditas pangan.

TABEL 1.
SAMPLE HARGA KOMODITAS PANGAN

Kode Provinsi	Provinsi	Komoditas	Tahun	Bulan	Harga
11	Aceh	Beras Premium	2021	Januari	12.072
12	Sumatera Utara	Beras Premium	2021	Januari	12.523
13	Sumatera Barat	Beras Premium	2021	Januari	13.096
14	Riau	Beras Premium	2021	Januari	14.605
15	Jambi	Beras Premium	2021	Januari	12.323
96	Papua Barat Daya	Daging Kerbau Segar (Lokal)	2025	April	0

Pra-pemrosesan data adalah tahap pembuka yang penting dalam melakukan pengolahan data, yang bertujuan memastikan bahwa dataset dalam kondisi bersih, konsisten, dan layak untuk dianalisis. Alur pra-pemrosesan data yang diimplementasikan dalam penelitian ini secara terstruktur

meliputi pembersihan nilai, seleksi dan reduksi fitur, transformasi data kategorikal, transformasi fitur bulan, fitur deret waktu, penanganan missing value, transformasi logaritma pada variabel target, hingga pembagian data berbasis temporal (temporal split). Dataset mentah awal yang diperoleh dari portal resmi Badan Pangan Nasional memiliki 27.722 baris dan 6 kolom fitur yang terdiri atas kode provinsi, nama provinsi, komoditas, tahun, bulan, dan target harga. Proses inspeksi data secara terstruktur berhasil mendeteksi adanya dua jenis anomali utama, yaitu 373 baris data yang memiliki nilai non-numerik atau cacat karakter struktural serta 79 baris data yang mencatat nilai Rp0 akibat ketiadaan pelaporan dari wilayah administrasi terkait. Untuk menjaga integritas sebaran data dan menghindari bias, seluruh baris data yang mengandung noise tersebut dihapus secara permanen dari dataframe utama, sehingga menghasilkan total 27.270 data bersih yang valid untuk digunakan ke tahap selanjutnya.

Tahap selanjutnya seluruh data diurutkan secara ketat berdasarkan urutan hierarki kronologis mulai dari variabel komoditas, nama provinsi, tahun, hingga bulan angka. Variabel bulan yang awalnya bertipe string alfabetis (Januari hingga Desember) dikonversi menjadi representasi numerik riil (1 hingga 12) guna membentuk urutan temporal sekuensial yang valid. Selain itu, fitur kategorikal nominal seperti nama provinsi dan komoditas ditransformasikan menjadi bentuk numerik terurut menggunakan teknik Label Encoding berbasis pemetaan kamus (dictionary mapping) untuk mempertahankan konsistensi indeks data.

Untuk memetakan efek ketergantungan waktu dan menangkap pola tren tanpa mengandalkan model statistik murni, dua fitur prediktor deret waktu utama dikonstruksi melalui fungsi pengelompokan (grouping) berbasis kombinasi entitas komoditas dan provinsi. Fitur pertama adalah harga lag 1 ($t-1$) yang digeser satu periode ke belakang untuk menangkap nilai harga aktual 1 bulan sebelumnya. Fitur kedua adalah harga moving average 3 ($t-1$, $t-2$, $t-3$) yang merekam rata-rata bergerak dari nilai harga pada 3 bulan terakhir sebagai indikator tren pergerakan harga jangka pendek. Proses pergeseran baris saat pembentukan kedua fitur tersebut secara otomatis memicu munculnya nilai kosong (NaN) pada baris-baris awal di setiap kelompok komoditas dan provinsi. Penanganan missing values ini diselesaikan melalui kombinasi dua tahap, yakni pengisian nilai kosong menggunakan nilai valid terdekat setelahnya (backfill) yang berjalan secara lokal di dalam grup masing-masing, dilanjutkan dengan median imputation global menggunakan nilai median umum dataset apabila masih terdapat sisa sel kosong pada awal urutan historis karena sifat median yang robust terhadap pengaruh outliers data [10].

Ketimpangan skala nominal harga antar-komoditas pangan di Indonesia sangat tinggi, seperti terlihat pada perbedaan rentang harga per kilogram yang sangat ekstrem antara komoditas beras dengan daging sapi. Kondisi ini dapat menyebabkan fungsi kerugian (loss function) algoritma regresi didominasi oleh komoditas berharga nominal tinggi.

Oleh karena itu, untuk mereduksi efek heteroskedastisitas dan menstabilkan varians skala target harga [11], diterapkan transformasi logaritma natural menggunakan persamaan (1) dan Seluruh proses pelatihan model, baik pada base learners maupun *Stacking Ensemble* dilakukan menggunakan data target yang telah ditransformasikan ke dalam skala logaritma. Setelah model menghasilkan prediksi, nilai prediksi tersebut dikembalikan ke skala harga asli menggunakan fungsi invers, sebagaimana ditunjukkan pada persamaan (2).

$$y_{\log} = \ln(\text{Harga} + 1) \quad (1)$$

$$\text{Harga}_{\text{pred}} = e^{y_{\log\text{-pred}}} - 1 \quad (2)$$

Dengan $Y_{\log}$ adalah nilai target baru dalam skala logaritma yang akan dipelajari oleh base learners dan meta-learner. In adalah Logaritma Natural. Harga adalah nilai nominal harga pangan asli dalam satuan Rupiah. $Y_{\log\text{-pred}}$ adalah nilai estimasi harga yang dikeluarkan oleh model *Stacking*. e adalah bilangan Euler. $\text{Harga}_{\text{pred}}$ adalah hasil prediksi final dalam satuan Rupiah yang siap dievaluasi menggunakan metrik RMSE dan MAE terhadap data aktual.

Pada tahap akhir pra-pemrosesan, dilakukan pembagian data latih dan data uji. penelitian ini menerapkan pendekatan temporal split sekuensial kronologis dengan rasio berimbang 80:20 [12]. Melalui pendekatan ini, data dipotong secara berdasarkan garis batas sekuensial waktu pergerakannya. Data latih mengambil 80% data yang mencakup rentang waktu periode tahun 2021 hingga pertengahan tahun 2024 dengan total 21.816 baris data. Data uji mengambil 20% yang mencakup periode akhir tahun 2024 hingga awal tahun 2025 dengan total 5.454 baris data. pembagian data dilakukan untuk menjaga keseimbangan distribusi kelas pada data latih dan uji [13].

Penelitian ini menerapkan model *Stacking Ensemble* untuk memprediksi harga komoditas pangan. Proses awal *Stacking Ensemble* adalah melatih empat *base learners* yaitu, *Ridge Regression*, *AdaBoost Regressor*, *Gradient Boosting Regressor*, dan *Extra Trees Regressor* secara terpisah pada data latih untuk menghasilkan prediksi. Hasil prediksi yang dihasilkan oleh setiap base learner selanjutnya disatukan untuk membentuk *meta-features*, yaitu fitur baru yang menampilkan performa awal dari masing-masing model terhadap data pelatihan. [14]. Komponen *Ridge Regression* disisipkan sebagai perwakilan model linier konvensional sekaligus untuk membantu meng-ekstrapolasi tren inflasi waktu makro secara internal di dalam sistem ensemble.

Kemudian *meta-learners* berupa *Decision Tree Regressor* digunakan untuk menghasilkan prediksi akhir yang menggabungkan kekuatan model dasar. *Decision Tree Regressor* dipilih dengan batasan regularisasi ketat ($\text{max_depth}=5, \text{min_samples_leaf}=5, \text{min_samples_split}=10$) melalui validasi silang internal 10-Fold ($\text{cv}=10$). Justifikasi utama penggunaan *Decision Tree* sebagai penggabung akhir adalah kemampuannya yang sangat fleksibel dalam

memetakan batas keputusan non-linier dari meta-features yang dihasilkan oleh model dasar[15].

Setelah proses *Stacking Ensemble* dilakukan evaluasi performa model menggunakan metrik RMSE dan koefisien determinasi R^2 pada data uji untuk membandingkan antara base model dan model *Stacking Ensemble* dianalisis untuk melihat peningkatan akurasi[16]. Rumus perhitungan untuk RMSE dan R^2 masing-masing ditulis pada persamaan (3), (4) dan (5).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

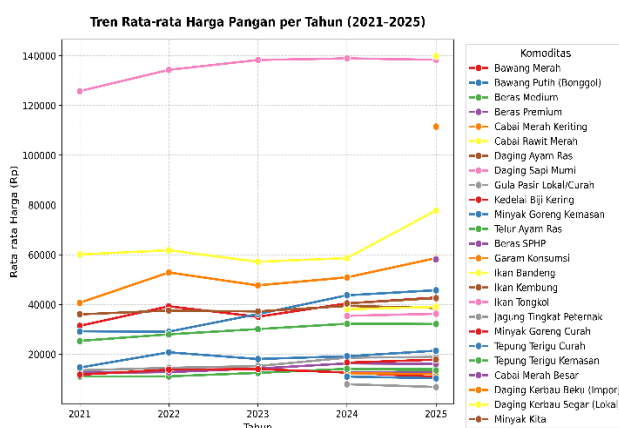
$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

Dengan R^2 digunakan sebagai mengukur kemampuan model dalam menjelaskan variasi data target, RMSE digunakan sebagai mengukur rata-rata kesalahan prediksi model terhadap nilai aktual, MAE digunakan untuk mengukur rata-rata kesalahan absolut prediksi model secara linier tanpa memberikan bobot lebih pada kesalahan besar, y_i merupakan nilai aktual, $\hat{y}_i$ merupakan nilai hasil prediksi model, $\bar{y}$ merupakan rata-rata dari seluruh nilai aktual, dan n merupakan jumlah data yang digunakan dalam evaluasi model.

III. HASIL DAN PEMBAHASAN

A. Analisis Deskriptif Data Harga Komoditas Pangan

Sebelum dilakukan proses pra-pemrosesan dan pemodelan menggunakan algoritma machine learning, dilakukan analisis deskriptif terhadap data harga komoditas pangan untuk memahami pola umum dan tren pergerakan harga selama periode penelitian. Analisis ini bertujuan untuk memberikan gambaran awal mengenai karakteristik data serta dinamika harga komoditas pangan di Indonesia.



Gambar 2. Tren Rata-rata Harga Pangan per Tahun 2021-2025

Gambar 2 menunjukkan tren rata-rata harga komoditas pangan nasional selama periode 2021 sampai 2025. Rata-rata harga dihitung berdasarkan seluruh komoditas dan provinsi yang terdapat dalam dataset. Dari visualisasi tersebut terlihat bahwa harga komoditas pangan mengalami peningkatan yang cukup signifikan dari tahun 2021 ke 2022, kemudian relatif stabil pada tahun 2023. Pada tahun 2024 terjadi penurunan harga yang cukup tajam, sebelum kembali mengalami kenaikan pada tahun 2025. Pola fluktuasi harga tersebut mengindikasikan bahwa pergerakan harga komoditas pangan tidak bersifat linier dan dipengaruhi oleh berbagai faktor eksternal, seperti kondisi pasokan dan distribusi, permintaan musiman, serta kebijakan pemerintah terkait stabilisasi harga pangan[17]. Fluktuasi harga yang kompleks dan tidak konsisten antar periode tersebut menunjukkan bahwa pendekatan prediksi sederhana berpotensi menghasilkan tingkat kesalahan yang tinggi. Oleh karena itu, diperlukan model prediksi yang mampu menangkap pola non-linear dan hubungan kompleks dalam data harga. Hal inilah yang menjadi dasar penggunaan pendekatan *machine learning* berbasis ensemble, khususnya metode *Stacking Ensemble* pada penelitian ini.

B. Hasil Performa Model

Setelah menerapkan pembagian data berbasis urutan sekuensial waktu (temporal split) dan transformasi logaritma pada variabel target, hasil pengujian performa prediksi pada data uji nominal menunjukkan kontras yang signifikan antar-algoritma[18]. Hasil komparasi performa evaluasi secara menyeluruh disajikan pada Tabel 2.

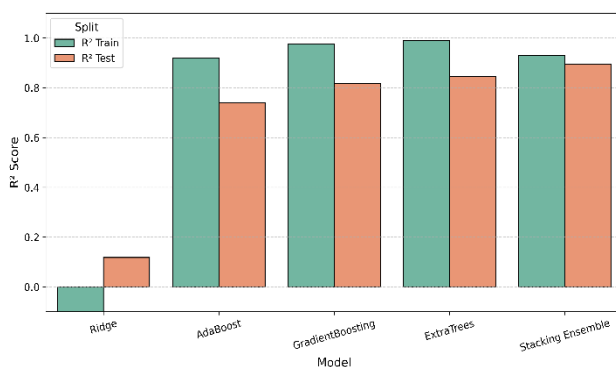
TABEL 2.
HASIL EVALUASI

Model	R^2 Latih	R^2 Uji	RMSE	MAE
Ridge Regression	-11.748	0.117	7.362	5.445
AdaBoost Regressor	0.919	0.739	4.003	2.959
Gradient Boosting Regressor	0.975	0.816	3.355	2.054
Extra Trees Regressor	0.988	0.844	3.091	1.982
Stacking Ensemble	0.930	0.895	2.531	1.461

Berdasarkan data pada Tabel 2, terlihat keterbatasan penggunaan model linier konvensional secara tunggal pada data historis pangan. Model Ridge Regression murni menghasilkan performa yang sangat buruk pada data latih dengan nilai R^2 Latih sebesar -11.7485 dan tingkat kesalahan MAE mencapai 5.445. Penurunan drastis ini terjadi karena Ridge Regression memaksakan asumsi hubungan garis lurus pada fitur input kategorikal terindeks serta target harga pangan yang sejatinya bergerak secara fluktuatif dan non-linier kompleks lintas wilayah. Model linier tunggal terbukti

tidak memiliki kapasitas yang cukup untuk mengadaptasi varians tabular yang sangat dinamis ini secara mandiri[19].

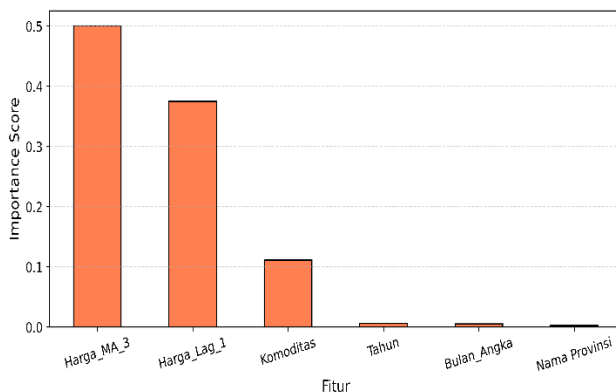
Di sisi lain, model individu berbasis pohon tunggal Gradient Boosting dan Extra Trees memperlihatkan performa yang tinggi di data latih namun mengalami gejala overfitting saat dihadapkan pada data masa depan di rentang data uji. Gradient Boosting dan Extra Trees nilai R^2 Latih yang sangat tinggi, namun menyusut pada R^2 Uji. Overfitting ini berhasil diatasi secara optimal oleh arsitektur Stacking Ensemble usulan. Model Stacking sukses mengamankan nilai R^2 Uji tertinggi sebesar 0.895, dengan nilai akumulasi error terkecil yang direpresentasikan oleh RMSE terendah sebesar 2.531 serta MAE sebesar 1.461. Untuk memvisualisasikan ketahanan model terhadap gejala overfitting, perbandingan R^2 Latih dan R^2 Uji disajikan pada Gambar 3.



Gambar 3. perbandingan R2 Latih dan R2 Uji

C. Feature Importance, Analisis Performa Per Komoditas, dan Distribusi Prediksi

Analisis Feature Importance dibuat melalui bobot internal dari model Extra Trees untuk melihat kontribusi masing-masing variabel terhadap keputusan prediksi harga. Hasil visualisasi skor kontribusi disajikan dalam Gambar 4.



Gambar 4. Feature Importance

Gambar 4, mengidentifikasi kekuatan prediktif utama yang didominasi oleh variabel-variabel transformasi temporal. Fitur rata-rata bergerak tiga bulanan (Harga_MA_3) memegang pengaruh tertinggi mutlak mencapai 50.06%,

disusul rapat oleh fitur lag satu bulan (Harga_Lag_1) sebesar 37.49%. Temuan empiris ini mengindikasikan struktur autoregressive yang sangat kuat pada pasar pangan domestik, di mana pergerakan harga saat ini sangat terikat oleh memori historis harga jangka pendek. Sementara itu, komponen kategoris makro seperti jenis Komoditas menyumbang 11.12%, sedangkan parameter Tahun, Bulan_Angka, dan Nama Provinsi berada di bawah 1%. Hal ini membuktikan bahwa variasi harga antar wilayah secara individual tidak lebih sensitif dibanding tren harga dari waktu ke waktu pada wilayah tersebut.

Pencapaian model *Stacking Ensemble* terletak pada kemampuannya mengendalikan bias dan varians, yang dibuktikan dengan perolehan Gap R^2 terkecil yaitu hanya sebesar 0.035. Hasil ini membuktikan bahwa bias linier dari Ridge Regression berhasil dimanfaatkan secara oleh pohon *meta-learner* untuk mengeksploitasi tren makro seiring waktu, sementara varians non-linier lokal dikompensasi dengan baik oleh prediksi dari model-model berbasis pohon. Dukungan batasan regularisasi pada tingkat *meta-learner* berhasil menahan model agar tidak menghafal pola data masa lalu, sehingga menciptakan model prediksi yang sangat kokoh untuk data temporal baru.

Untuk melihat stabilitas dari interaksi fitur tersebut diuji lebih mendalam melalui analisis dekomposisi performa prediksi *Stacking Ensemble* pada masing-masing jenis komoditas pangan. Berikut adalah performa stacking per komoditas ditunjukkan pada table 3.

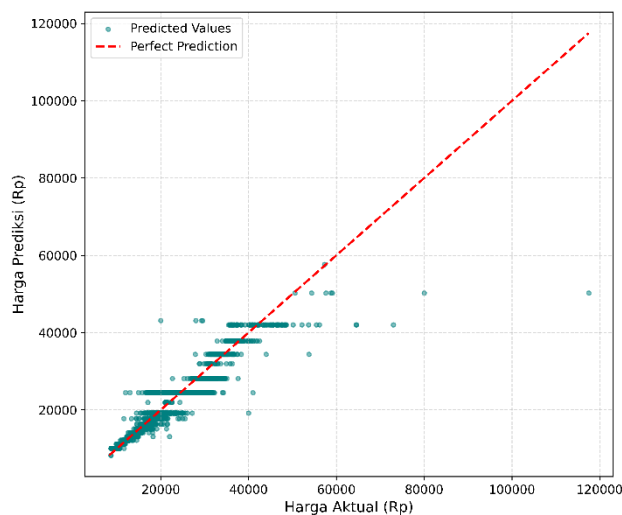
TABEL 3.
PERFORMA STACKING PER KOMODITAS

Komoditas	Data	R2	RMSE	MAE
Tepung Terigu Kemasan	603	0,835	1.180	415
Minyak Goreng Kemasan	1.825	0,829	1.595	994
Tepung Terigu Curah	600	0,802	874	489
Telur Ayam Ras	1.828	0,579	3.769	2.720
Minyak Goreng Curah	488	0,445	1.320	699
Minyak Kita	110	-36,463	4.454	2.675

Hasil pengujian spesifik per komoditas pada Tabel 3, menunjukkan reliabilitas tinggi pada komoditas tepung terigu kemasan $R^2 = 0.835$ dan minyak goreng kemasan $R^2 = 0.829$. Namun, akurasi mengalami penurunan pada komoditas dengan volatilitas tinggi seperti Telur Ayam Ras $R^2 = 0.579$ dan Minyak Goreng Curah $R^2 = 0.445$.

Anomali terbesar ditemukan pada komoditas "Minyak Kita" yang mencatat nilai R^2 negatif ekstrem -36,463. Hal ini terjadi karena keterbatasan volume data historis karena N

Data hanya 110 serta sifat komoditas tersebut sebagai program subsidi pemerintah yang harganya dipatok secara regulasi (Harga Eceran Tertinggi/HET) baru pada pertengahan linimasa data, sehingga polanya tidak dapat direkonstruksi dengan baik oleh fitur Lag maupun Moving Average berbasis tren pasar bebas komoditas lainnya.



Gambar 5. Grafik Harga Actual dan Harga Prediksi

Visualisasi pada Gambar 4 memberikan gambaran yang lebih jelas mengenai kinerja model. Semakin dekat titik prediksi terhadap garis tersebut, semakin kecil error yang dihasilkan model. Sebaran titik prediksi menunjukkan bahwa sebagian besar nilai prediksi berada sangat dekat dengan garis referensi prediksi sempurna. Dengan demikian, ini menegaskan bahwa kemampuan *Stacking Ensemble* lebih baik dibandingkan model tunggal, sekaligus menunjukkan stabilitas model pada rentang harga yang luas[20].

Keunggulan model *Stacking Ensemble* dibandingkan model tunggal pada penelitian ini disebabkan oleh kemampuannya dalam mengombinasikan karakteristik dan kelebihan dari masing-masing *base learners*. *AdaBoost Regressor* efektif dalam memperbaiki kesalahan prediksi secara iteratif, *Gradient Boosting Regressor* unggul dalam menangkap pola non-linear yang kompleks, sementara *Extra Trees Regressor* mampu mengurangi variansi model melalui proses randomisasi yang tinggi. *Ridge Regression* sebagai *base learner* linier memberikan kontribusi kompensasi bias yang krusial bagi ekosistem ensemble. Meskipun performa mandiri dari model *Ridge Regression* bernilai negatif akibat ketidakmampuannya memetakan data kategorikal terindeks dan target yang fluktuatif secara mandiri, bias linier yang dimilikinya justru dimanfaatkan oleh arsitektur *Stacking* untuk memetakan tren makro (seperti kecenderungan inflasi jangka panjang berbasis fitur runtun waktu). Dengan menggabungkan keluaran ketiga model tersebut sebagai *meta-features*, *Decision Tree Regressor* selaku *meta-learner* dapat mempelajari pola kesalahan dan hubungan residual antar model dasar secara adaptif. Proses ini memungkinkan

Stacking Ensemble menghasilkan prediksi yang lebih stabil dan robust terhadap fluktuasi harga ekstrem, serta meminimalkan bias dan variansi yang umumnya muncul pada model tunggal[15]. Oleh karena itu, Metode *Stacking Ensemble* tidak hanya mampu meningkatkan tingkat akurasi secara kuantitatif, tetapi juga memperlihatkan kestabilan performa pada berbagai rentang harga komoditas pangan. Hal tersebut terlihat dari hasil visualisasi prediksi yang cenderung mengikuti garis referensi ideal, sehingga menunjukkan bahwa model memiliki kemampuan prediksi yang konsisten dan mendekati nilai aktual.

IV. KESIMPULAN

Penelitian ini berhasil membuktikan bahwa rekonstruksi model *Stacking Ensemble Learning* yang diperkuat dengan fitur komponen deret waktu yaitu Lag dan Moving Average serta transformasi logaritma target mampu menyajikan performa prediksi harga pangan multi-provinsi yang jauh lebih realistis, stabil, dan bebas dari risiko kebocoran data temporal. Melalui pembagian data berbasis temporal split, model *Stacking Ensemble* usulan terbukti menjadi arsitektur terbaik dengan performa R^2 Uji sebesar 0.895, RMSE sebesar 2.531, MAE sebesar 1.461, Model ini teruji sangat kokoh dalam mengendalikan bias dan varians generalisasi dengan menyusutkan jarak celah overfitting hingga menyentuh angka 0.035. Secara kuantitatif, model *Stacking Ensemble* yang diusulkan menunjukkan performa superior dibandingkan seluruh model tunggal. Dengan memanfaatkan *meta-learner* *Decision Tree Regressor* tingkat kedua, model ini mampu mengombinasikan karakteristik komplementer secara adaptif mengompensasi bias linier makro dari *Ridge Regression* serta mereduksi varians lokal non-linier dari *AdaBoost*, *Gradient Boosting*, dan *Extra Trees*.

Pergerakan harga pangan nasional didominasi oleh efek inersia histori jangka pendek sebesar 87.55%, memberikan petunjuk taktis bagi instansi pemerintah untuk melakukan langkah intervensi pasar secara agresif dalam kurun waktu kurang dari satu bulan guna memitigasi risiko inflasi berkelanjutan. Keterbatasan model ini ditemukan pada komoditas bersubsidi pemerintah seperti Minyak Kita akibat adanya distorsi data horizontal oleh kebijakan HET, yang disarankan menjadi fokus riset lanjutan menggunakan pendekatan pemodelan intervensi non-linier khusus seperti indikator cuaca ekstrem, volume pasokan gudang, serta variabel kuantitatif intervensi kebijakan guna meningkatkan fleksibilitas model terhadap anomali struktural.

DAFTAR PUSTAKA

- [1] R. A. Chintia, R. Destiningsih, P. Studi, E. Pembangunan, U. Tidar, and K. Magelang, "Pengaruh harga komoditas pangan terhadap inflasi di kota semarang," vol. 27, no. 2, pp. 244–258.
- [2] O. Imandira, R. Farisi, N. Jannah, and R. Insania, "Prediksi Harga Komoditas Pangan di Indonesia Menggunakan Backpropagation," vol. 3, no. 1, pp. 91–101, 2022.
- [3] P. S. Agribisnis and U. J. Soedirman, "Pengaruh Fluktuasi Harga Cabai Rawit Merah Terhadap Inflasi di Kabupaten Banyumas," vol. 10, no. 2020, pp. 1866–1877, 2024.

- [4] A. R. Pratama, F. Wabula, H. Ilmandry, M. L. Isabela, and R. Sianipar, "Literature Review The Impact of Machine Learning in Modern Industries," no. 2021, 2025.
- [5] O. Sagi and L. Rokach, "Ensemble learning : A survey," no. January, pp. 1–18, 2018, doi: 10.1002/widm.1249.
- [6] A. N. Haya and M. Y. Ramme, "Penerapan Algoritma Stacking Ensemble Machine Learning Berbasis Pohon untuk Prediksi Penyakit Diabetes," vol. 2024, no. Senada, pp. 954–961, 2024.
- [7] I. M. Tirta, A. Riski, and N. Sholikhah, "Prediksi Harga Saham PT Bank Rakyat Indonesia Tbk Menggunakan AUTOML H2O Pendahuluan," vol. 23, no. September, pp. 397–404, 2024.
- [8] M. N. Hibatulloh, G. D. Prakoso, A. Dwi, P. Yunus, and T. D. Putra, "Prediksi Harga Rumah di Bandung 2024 Menggunakan Ensemble Learning : Analisis Komparatif dan Interpretabilitas," vol. 10, no. 2, pp. 484–493, 2025, doi: 10.30591/jpit.v9ix.xxx.
- [9] Badan Pangan Nasional, "Rata-rata Harga Pangan Bulanan Tingkat Konsumen Provinsi 2025." [Online]. Available: <https://data.badanpangan.go.id/datasetpublications/tq2/harga-pangan-tingkat-konsumen-provinsi>
- [10] J. Ilmu, T. Informasi, V. No, and J. Hal, "STUDI PERBANDINGAN ALGORITMA MACHINE LEARNING," vol. 1, no. 1, pp. 22–28, 2026.
- [11] A. Karim, B. Bangun, S. Prayetno, and M. Afrendi, "Optimasi Prediksi Harga Sawit Menggunakan Teknik Stacking Algoritma Machine Learning dan Deep Learning dengan SMOTE," vol. 7, no. 1, pp. 638–645, 2025, doi: 10.47065/bits.v7i1.7239.
- [12] K. Zhang, R. Xia, Y. Wang, Y. Chen, X. Wang, and J. Dou, "Stack Coupling Machine Learning Model Could Enhance the Accuracy in Short-Term Water Quality Prediction," pp. 1–29, 2025.
- [13] A. Agustina, T. Tukino, B. Huda, and E. Novalia, "Prediksi Volume Penjualan Gadget Berdasarkan Promo dan Channel Penjualan Menggunakan Random Forest," vol. 4, no. 1, pp. 85–91, 2025.
- [14] G. G. Ghiffary, N. T. Amanda, R. Ardhani, B. Sartono, A. Rizki, and S. S. Data, "Analisis Kinerja Model Stacking Berbasis Random Forest dan SVM dalam Klasifikasi Rumah Tangga Berdasarkan Garis Kemiskinan Makanan di Provinsi Jawa Barat," vol. 5, no. 3, pp. 2244–2265, 2024.
- [15] F. R. Salwa, M. Novita, and R. Renaldy, "Peningkatan Performa Prediksi Survival Pasien Gagal Jantung Menggunakan Stacking Ensemble Learning Improving Survival Prediction Performance of Heart Failure Patients Using Stacking Ensemble Learning," no. 3, pp. 614–625, 2025, doi: 10.26798/jiko.v9i3.2126.
- [16] R. A. Saputra, A. Pratama, S. Informasi, U. Nahdlatul, and U. Kalimantan, "Implementasi Decision Tree Untuk Prediksi Harga Rumah Di Daerah Tebet," vol. 6, no. 2, pp. 4–10, 2025.
- [17] I. Marina, D. Sukmawati, E. Juliana, and N. Safa, "Dinamika Pasar Komoditas Pangan Strategis : Analisis Fluktuasi Harga Dan Produksi," vol. 12, no. 1, pp. 160–168, 2024.
- [18] U. Royal, "Stacking Ensemble Model Machine Learning Deteksi Dini Risiko Kesehatan Mental Di Lingkungan Pendidikan," vol. 4307, no. August, pp. 4256–4266, 2024.
- [19] A. F. Nugraha, R. Faticha, A. Aziza, and Y. Pristyanto, "Penerapan metode Stacking dan Random Forest untuk Meningkatkan Kinerja Klasifikasi pada Proses Deteksi Web Phishing," vol. 7, no. 1, 2022.
- [20] A. S. Alfath and A. K. Wardhana, "Hypertension Risk Prediction Using Stacking Ensemble of CatBoost , XGBoost , and LightGBM : A Machine Learning Approach," vol. 9, no. 6, pp. 3146–3156, 2025.