

PENERAPAN ALGORITMA XGBOOST (*EXTREME GRADIENT BOOSTING*) UNTUK PREDIKSI INFLASI PANGAN SUMATERA UTARA

Eka Yulia Putri, Arnita, Said Iskandar Al Idrus, Zulfahmi Indra, Chairunisah

Ilmu Komputer, Universitas Negeri Medan
Jalan Willem Iskandar Pasar V Medan Estate
ekayuliaputri33@gmail.com

ABSTRAK

Inflasi pangan merupakan salah satu komponen utama yang memengaruhi stabilitas ekonomi daerah, terutama di Provinsi Sumatera Utara yang menunjukkan fluktuasi relatif tinggi dalam beberapa tahun terakhir. Kondisi ini menuntut adanya metode peramalan yang akurat dan andal sebagai dasar pengambilan kebijakan. Penelitian ini bertujuan untuk menerapkan algoritma *Extreme Gradient Boosting* (XGBoost) dalam memprediksi inflasi pangan di Provinsi Sumatera Utara. Data yang digunakan mencakup harga harian sepuluh komoditas pangan strategis yang diperoleh dari PIHPS serta data inflasi subkelompok makanan dari BPS selama periode Januari 2021 hingga Juli 2025. Model dikembangkan melalui proses *hyperparameter tuning* untuk memperoleh performa optimal, dan dievaluasi menggunakan metrik *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), serta koefisien determinasi (R^2). Hasil penelitian menunjukkan bahwa model XGBoost memiliki kinerja yang baik dengan nilai RMSE sebesar 0,5225, MAE sebesar 0,4530, dan R^2 mencapai 89%. Selain itu, hasil peramalan menunjukkan pola inflasi yang cenderung fluktuatif namun tetap stabil dalam jangka pendek. Analisis menggunakan metode SHAP mengungkapkan bahwa variabel perubahan indeks harga pangan (*all_index_pctchg*) merupakan faktor paling dominan dalam memengaruhi inflasi. Dengan demikian, XGBoost terbukti efektif dalam memodelkan dan memprediksi inflasi pangan sebagai dasar perumusan kebijakan ekonomi daerah.

Kata kunci : Inflasi Pangan, XGBoost, Machine Learning, Forecasting, Time Series, Sumatera Utara

1. PENDAHULUAN

Inflasi merupakan indikator makroekonomi yang mencerminkan kenaikan harga barang dan jasa secara umum dan berkelanjutan. Stabilitas inflasi berperan penting dalam menjaga daya beli masyarakat, pertumbuhan ekonomi, serta keberlanjutan pembangunan [1].

Di Indonesia, pengendalian inflasi menjadi prioritas Bank Indonesia dengan target inflasi sebesar $3\% \pm 1\%$ [2].

Komoditas pangan merupakan penyumbang terbesar inflasi di Indonesia, dengan kontribusi 42,7% dari keranjang Indeks Harga Konsumen [3].

Fluktuasi harga pangan memiliki dampak langsung terhadap kesejahteraan masyarakat, terutama kelompok berpendapatan rendah, serta berpotensi menimbulkan tekanan inflasi yang bersifat volatile. Oleh karena itu, pengendalian inflasi pangan menjadi aspek krusial dalam stabilitas ekonomi.[4].

Pada tingkat daerah, dinamika inflasi menunjukkan karakteristik yang beragam. Yangmana perbedaan struktur ekonomi, kondisi infrastruktur, serta sistem distribusi menyebabkan variasi pola inflasi antar wilayah [5].

Provinsi Sumatera Utara merupakan salah satu daerah dengan tingkat inflasi yang relatif fluktuatif dan cenderung lebih tinggi dibandingkan rata-rata nasional. Kelompok makanan, minuman, dan tembakau menjadi kontributor utama inflasi, dengan komoditas strategis seperti cabai merah dan daging ayam ras yang sering memicu lonjakan harga [6][7][8].

Kondisi ini mencerminkan adanya ketidakpastian dalam pergerakan harga pangan yang perlu diantisipasi melalui pendekatan analisis yang tepat.

Permasalahan utama dalam pengendalian inflasi pangan terletak pada karakteristik data yang dinamis, nonlinier, serta dipengaruhi oleh faktor musiman dan keterkaitan antarvariabel yang kompleks. Metode peramalan konvensional sering kali memiliki keterbatasan dalam menangkap pola tersebut secara optimal [9][10]. Selain itu, penelitian terkait peramalan inflasi pa

ngan di tingkat daerah, khususnya di Provinsi Sumatera Utara, masih terbatas dan umumnya menggunakan pendekatan linier yang kurang adaptif terhadap kompleksitas data [11][12].

Seiring dengan perkembangan teknologi, pendekatan machine learning menjadi alternatif yang semakin banyak digunakan dalam analisis data ekonomi karena kemampuannya dalam mengolah data berskala besar dan menangkap hubungan kompleks antarvariabel [13].

Salah satu algoritma yang memiliki performa baik dalam berbagai kasus prediksi adalah *Extreme Gradient Boosting* (XGBoost), yaitu metode ensemble berbasis *decision tree* yang bekerja secara iteratif untuk meminimalkan kesalahan prediksi [14][15].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis pola pergerakan inflasi pangan, mengidentifikasi pengaruh harga komoditas pangan strategis, efek lag inflasi, serta faktor musiman terhadap inflasi pangan, serta membangun model peramalan inflasi pangan di Provinsi Sumatera Utara menggunakan algoritma

XGBoost. Model yang dikembangkan akan melalui proses hyperparameter tuning untuk memperoleh performa optimal dan dievaluasi menggunakan metrik *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), serta koefisien determinasi (R^2). Hasil penelitian ini diharapkan dapat memberikan model prediksi yang akurat dan menjadi dasar dalam perencanaan serta pengambilan kebijakan pengendalian inflasi pangan di tingkat daerah.

2. TINJAUAN PUSTAKA

2.1. Inflasi Pangan

Inflasi pangan didefinisikan sebagai kenaikan rata-rata harga komoditas pangan yang diukur melalui perubahan Indeks Harga Konsumen (IHK) kelompok makanan dari waktu ke waktu. Di Indonesia, penghitungan inflasi pangan dilakukan oleh Badan Pusat Statistik (BPS) dengan mencakup berbagai komoditas utama seperti beras, cabai, bawang, daging ayam, telur, serta komoditas strategis lainnya [16].

Dalam kerangka pengendalian inflasi, inflasi pangan dikategorikan sebagai *volatile food* karena sangat dipengaruhi oleh faktor musiman, gangguan pasokan, serta fluktuasi harga komoditas baik domestik maupun global [17].

Mengingat besarnya proporsi pengeluaran rumah tangga untuk kebutuhan pangan, stabilitas harga pangan menjadi faktor penting dalam pengendalian inflasi nasional [18].

Secara konseptual, perhitungan IHK pangan mengikuti prinsip indeks Laspeyres berbobot tetap, di mana setiap komoditas memiliki bobot yang mencerminkan proporsi pengeluaran rumah tangga [17].

Komoditas dengan bobot yang lebih besar akan memberikan pengaruh yang lebih signifikan terhadap perubahan indeks. Secara matematis, IHK pangan dapat dirumuskan sebagai berikut:

$$IHK_{pangan,t} = \sum_{i=1}^n \omega_i \cdot \frac{p_{i,t}}{p_{i,0}} \times 100 \quad (1)$$

dengan $IHK_{pangan,t}$ menyatakan indeks harga konsumen pangan pada periode t , $p_{i,t}$ merupakan harga komoditas ke- i pada periode t , $p_{i,0}$ merupakan harga komoditas ke- i pada periode dasar, ω_i adalah bobot pengeluaran untuk komoditas ke- i , dan n adalah jumlah komoditas dalam keranjang konsumsi.

Bobot tersebut mencerminkan tingkat kontribusi masing-masing komoditas terhadap inflasi pangan, sehingga semakin besar bobot suatu komoditas, semakin besar pula pengaruhnya terhadap perubahan indeks harga.

2.2. Peramalan

Peramalan (*forecasting*) merupakan proses memprediksi kondisi di masa mendatang berdasarkan data historis dan informasi saat ini melalui analisis pola dan tren [19].

Meskipun tidak menghasilkan prediksi yang pasti, peramalan bertujuan memberikan estimasi yang

mendekati kondisi aktual sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan.

Dalam konteks penelitian ini, peramalan digunakan untuk memprediksi tingkat inflasi pangan di Provinsi Sumatera Utara dengan memanfaatkan data historis harga komoditas pangan. Pendekatan peramalan berbasis data runtun waktu (*time series*) digunakan dengan asumsi bahwa pola historis data, seperti tren dan fluktuasi, akan berlanjut pada periode berikutnya [9][18].

2.3. Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) merupakan salah satu algoritma ensemble learning berbasis *decision tree* yang bekerja secara iteratif dengan menggabungkan sejumlah *weak learners* untuk menghasilkan model prediktif yang kuat. Metode ini pertama kali dikembangkan dari konsep gradient boosting oleh Friedman (2001), yang membangun model secara bertahap (*additive modeling*) dengan meminimalkan fungsi kerugian (*loss function*) pada setiap iterasi [21].

XGBoost, yang diperkenalkan oleh Chen dan Guestrin (2016), merupakan pengembangan dari metode gradient boosting dengan berbagai optimasi, seperti penggunaan turunan orde dua (gradien dan Hessian), regularisasi untuk mengontrol kompleksitas model, serta efisiensi komputasi yang tinggi. Hal ini memungkinkan XGBoost menghasilkan performa yang lebih baik dibandingkan metode boosting konvensional [14].

Secara matematis, model XGBoost dinyatakan sebagai penjumlahan dari sejumlah fungsi pohon keputusan sebagai berikut:

$$\hat{y}_i^{(t)} = \sum_{k=1}^K f_k(x_i) \quad (2)$$

di mana $f_k(x_i)$ merupakan fungsi prediksi dari pohon ke- k . Model dibangun secara iteratif, di mana setiap pohon baru berperan untuk memperbaiki kesalahan dari model sebelumnya.

Untuk mengoptimalkan model, XGBoost menggunakan fungsi objektif yang terdiri dari fungsi *loss* dan regularisasi:

$$Obj^{(t)} = \sum_{i=1}^t l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^t \Omega(f_k) \quad (3)$$

Fungsi *loss* mengukur kesalahan prediksi terhadap nilai aktual, sedangkan fungsi regularisasi digunakan untuk mengontrol kompleksitas model agar tidak terjadi *overfitting*. Bentuk fungsi regularisasi dinyatakan sebagai:

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (4)$$

dengan T adalah jumlah daun (*leaf*), w_j adalah bobot pada daun ke- j , serta γ dan λ merupakan parameter regularisasi.

Dalam proses pembentukan pohon, XGBoost menggunakan nilai *gain* sebagai kriteria untuk menentukan pemisahan node. Nilai *gain* dirumuskan sebagai:

$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (5)$$

Pemisahan node dilakukan apabila nilai *gain* bernilai positif, yang menunjukkan peningkatan kualitas model. Secara keseluruhan, XGBoost memiliki keunggulan dalam menangani data nonlinier, mengontrol kompleksitas model melalui regularisasi, serta memberikan kinerja prediksi yang tinggi. Oleh karena itu, algoritma ini banyak digunakan dalam berbagai penelitian, termasuk dalam analisis dan prediksi inflasi pangan.

2.4. Hyperparameter

Hyperparameter merupakan parameter yang ditentukan sebelum proses pelatihan model dan berfungsi untuk mengontrol proses pembelajaran serta kompleksitas model yang dihasilkan [22].

Dalam algoritma XGBoost, hyperparameter seperti *learning_rate*, *max_depth*, *n_estimators*, *subsample*, serta parameter *regularisasi* memiliki peran penting dalam menentukan kinerja model. Pemilihan nilai hyperparameter yang tepat sangat berpengaruh terhadap kemampuan model dalam menangkap pola data serta menghindari overfitting maupun underfitting.

Oleh karena itu, diperlukan proses hyperparameter tuning untuk menemukan kombinasi parameter yang optimal sehingga dapat meningkatkan akurasi model [23]. Proses ini menjadi tahap penting dalam pengembangan model machine learning, karena kinerja model tidak hanya ditentukan oleh algoritma yang digunakan, tetapi juga oleh konfigurasi parameter yang diterapkan [24].

2.5. SHAP (Shapley Additive Explanations)

SHAP (*Shapley Additive Explanations*) merupakan salah satu metode dalam *explainable artificial intelligence* (XAI) yang digunakan untuk menjelaskan kontribusi masing-masing variabel terhadap hasil prediksi model machine learning. Pendekatan ini didasarkan pada konsep nilai Shapley dalam game theory, yang membagi kontribusi setiap fitur secara adil terhadap output model [25].

Dalam SHAP, prediksi model tidak hanya dipandang sebagai nilai akhir, tetapi sebagai hasil agregasi kontribusi dari seluruh fitur yang berinteraksi secara simultan [26].

Dalam perkembangannya, SHAP banyak digunakan untuk menjelaskan model yang kompleks karena mampu memberikan interpretasi baik secara lokal maupun global tanpa mengorbankan akurasi.

2.6. Persentase Kesalahan

Untuk mengukur performa algoritma dalam peramalan, digunakan pengukuran nilai galat sebagai pendekatan untuk mengevaluasi perbedaan antara nilai prediksi dan nilai aktual. Pengukuran ini penting dalam menilai tingkat akurasi model serta menjadi dasar dalam meningkatkan kinerja model prediksi di berbagai bidang, seperti pembelajaran mesin, statistik, dan ekonomi [27].

RMSE digunakan untuk mengukur besarnya kesalahan prediksi dengan penalti lebih besar pada error besar, MAE untuk mengukur rata-rata kesalahan absolut yang lebih mudah diinterpretasikan, serta R^2 untuk mengukur kemampuan model dalam menjelaskan variasi data aktual. Evaluasi dilakukan pada data pengujian untuk memastikan kemampuan generalisasi model.

2.7. Penelitian Terdahulu

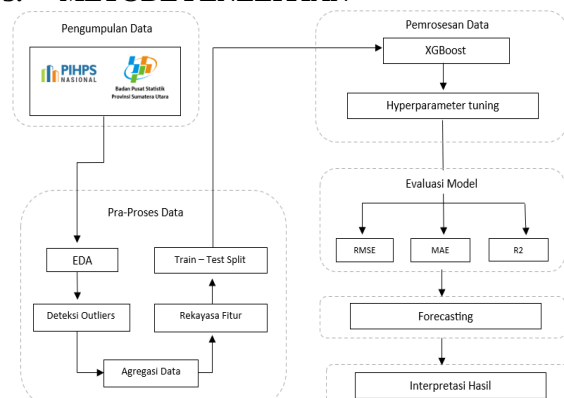
Beberapa penelitian terdahulu menunjukkan bahwa metode peramalan berbasis machine learning, khususnya XGBoost, memiliki kinerja yang baik dalam memodelkan data ekonomi yang bersifat kompleks dan nonlinier. Penelitian sebelumnya menunjukkan bahwa algoritma machine learning, termasuk XGBoost, mampu memprediksi tingkat inflasi dengan akurasi yang lebih tinggi dibandingkan metode konvensional serta efektif dalam menangani hubungan nonlinier dan dinamika data inflasi [28].

Studi lain menunjukkan bahwa XGBoost memiliki tingkat akurasi yang lebih baik dibandingkan metode deret waktu tradisional seperti ARIMA dalam memprediksi produksi komoditas pangan, yang mengindikasikan keunggulan metode ini dalam menangkap pola data yang kompleks [29]. Selain itu, penelitian yang memanfaatkan data harga komoditas pangan dan inflasi menunjukkan bahwa model XGBoost memberikan kinerja terbaik dibandingkan metode lain seperti SARIMA dan LSTM dalam prediksi inflasi pangan di Indonesia [30].

Penelitian lainnya juga menunjukkan bahwa XGBoost mampu menghasilkan prediksi dengan tingkat akurasi yang tinggi pada data harga pangan dan inflasi [31][32].

Secara keseluruhan, hasil penelitian terdahulu menunjukkan bahwa XGBoost merupakan metode yang efektif dan relevan untuk digunakan dalam prediksi inflasi pangan.

3. METODE PENELITIAN



Gambar 1. Diagram alur penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode peramalan deret waktu (*time series forecasting*) berbasis *machine learning* untuk memprediksi inflasi pangan di Provinsi

Sumatera Utara. Model yang digunakan adalah *Extreme Gradient Boosting* (XGBoost).

Proses penelitian dilakukan secara sistematis sesuai dengan alur penelitian yang mencakup tahapan pengumpulan data, pra-pemrosesan data, pembangunan dan optimasi model, evaluasi kinerja model, serta peramalan dan interpretasi hasil.

3.1. Pengumpulan Data

Penelitian ini menggunakan data sekunder yang diperoleh dari sumber resmi. Data harga sepuluh komoditas pangan strategis diperoleh dari Pusat Informasi Harga Pangan Strategis (PIHPS) Nasional (<https://www.bi.go.id/hargapangan>) dengan cakupan lima kota di Provinsi Sumatera Utara, yaitu Medan, Pematangsiantar, Padangsidimpuan, Sibolga, dan Gunungsitoli, dalam frekuensi harian.

Data inflasi subkelompok makanan diperoleh dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara dalam bentuk inflasi bulanan (month-to-month). Data periode 2021–2023 diperoleh melalui publikasi resmi, sedangkan periode 2024–2025 diperoleh melalui permohonan data ke BPS. Kemudian data bobot konsumsi komoditas juga digunakan sebagai variabel pendukung dalam pembentukan indeks harga pangan.

3.2. Pra-pemrosesan Data

Pra-pemrosesan data dilakukan untuk menyiapkan data sebelum pemodelan. Tahap ini diawali dengan *exploratory data analysis* (EDA) untuk memahami struktur data. Selanjutnya dilakukan pemeriksaan kualitas data, termasuk identifikasi *missing values* dan duplikasi. Penanganan *outlier* dilakukan menggunakan metode *Interquartile Range* (IQR) dan analisis visual, namun nilai ekstrem tetap dipertahankan karena mencerminkan kondisi pasar yang nyata. Setelah itu, data harga harian diagregasi menjadi harga bulanan dan digabungkan dengan data inflasi untuk membentuk dataset terpadu. Tahap berikutnya adalah perhitungan indeks harga pangan, baik per komoditas maupun indeks total berbobot. Kemudian dilakukan rekayasa fitur dengan membentuk variabel lag, perubahan indeks, persentase perubahan harga komoditas, serta variabel musiman. Terakhir, dataset dibagi menjadi data pelatihan (80%) dan pengujian (20%) menggunakan pendekatan *time-series split* untuk menjaga urutan temporal data.

3.3. Pemrosesan Data

Selanjutnya, pemodelan dilakukan menggunakan algoritma XGBoost. Tahap awal dimulai dengan membangun model baseline menggunakan parameter awal sebagai acuan performa. Setelah itu, dilakukan optimasi hyperparameter menggunakan *GridSearchCV* dengan pendekatan *TimeSeriesSplit* untuk memperoleh kombinasi parameter terbaik. Model hasil tuning kemudian digunakan sebagai model akhir dalam proses evaluasi dan peramalan inflasi pangan.

3.4. Evaluasi Model

Kemudian untuk menilai kinerja prediksi XGBoost dalam memodelkan inflasi pangan dilakukan evaluasi model. Penilaian dilakukan menggunakan tiga metrik, yaitu Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan koefisien determinasi (R^2).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (8)$$

di mana y_i merupakan nilai aktual, $\hat{y}_i$ adalah nilai prediksi, $\bar{y}$ adalah rata-rata nilai aktual, dan n adalah jumlah observasi. Evaluasi dilakukan pada data pengujian (*test set*) untuk memastikan kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

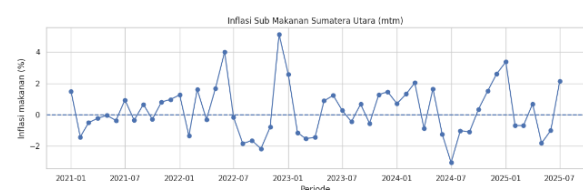
3.5. Peramalan (Forecasting)

Tahap akhir penelitian adalah melakukan peramalan inflasi pangan menggunakan model XGBoost terbaik yang diperoleh dari proses *hyperparameter tuning*. Metode yang digunakan adalah *multi-step autoregressive forecasting*, di mana hasil prediksi pada satu periode digunakan sebagai input untuk memprediksi periode berikutnya.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Penelitian ini menggunakan dua jenis data utama, yaitu data harga komoditas pangan harian sebanyak 58.555 observasi dan data inflasi subkelompok makanan bulanan sebanyak 55 observasi untuk periode 2021–2025. Data harga mencakup sepuluh komoditas strategis pada lima kota di Provinsi Sumatera Utara. Perbedaan frekuensi data menunjukkan perlunya proses agregasi sebelum analisis lanjutan.



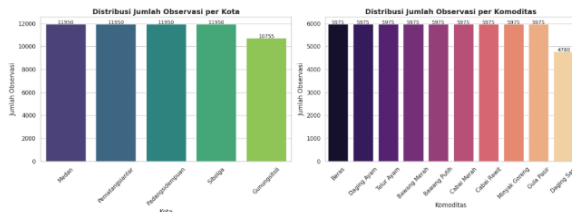
Gambar 2. Tren Inflasi Subkelompok Makanan Provinsi Sumatera Utara (2021–2025)

Tren inflasi menunjukkan pola fluktuatif dengan beberapa periode deflasi, yang mencerminkan dinamika harga pangan yang dipengaruhi oleh faktor musiman dan kondisi pasar.

4.2. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) merupakan tahap awal dalam analisis data yang bertujuan untuk memahami karakteristik dataset serta mengidentifikasi pola dan hubungan yang terdapat di dalam data [33]. Pada tahap ini dilakukan evaluasi struktur data dan distribusi observasi. Dataset terdiri atas lima variabel

utama, yaitu *date*, *kota*, *komoditas*, *harga*, dan *tahun_file*, yang telah sesuai untuk analisis deret waktu.



Gambar 3. Distribusi Observasi Harga Pangan per Kota dan Komoditas

Distribusi observasi menunjukkan bahwa data relatif merata pada sebagian besar kota dengan rata-rata sekitar 11.950 observasi, meskipun terdapat perbedaan pada beberapa wilayah seperti Gunungsitoli yang memiliki jumlah observasi lebih rendah. Pada tingkat komoditas, distribusi data juga relatif seimbang, meskipun daging sapi memiliki jumlah observasi yang lebih sedikit dibandingkan komoditas lainnya. Secara keseluruhan, distribusi data tidak menunjukkan ketimpangan yang signifikan sehingga layak digunakan dalam proses pemodelan.

4.3. Cek Missing Value

Tabel 1. Missing Value pada Dataset Harga Pangan

date	0
Kota	0
Komoditas	0
Harga	2276
Tahun_file	0

Nilai hilang pada variabel harga tidak disebabkan oleh kesalahan pencatatan, melainkan karena tidak tersedianya data pada tanggal tertentu. Oleh karena itu, tidak dilakukan imputasi, mengingat data akan diagregasi ke dalam bentuk bulanan sehingga nilai rata-rata tetap mampu merepresentasikan kondisi harga secara memadai. Selain itu, imputasi berpotensi mengganggu pola volatilitas harga yang bersifat alami.

4.4. Data Duplikat

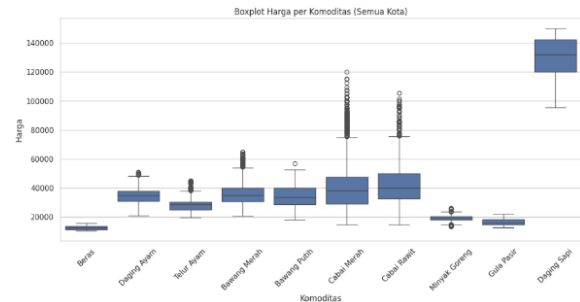
Selanjutnya dilakukan pemeriksaan terhadap keberadaan data duplikat untuk memastikan integritas dataset.

Tabel 2. Jumlah Data Duplikat

Jumlah Baris Duplikat	0
-----------------------	---

Berdasarkan hasil analisis, tidak ditemukan adanya baris data yang bersifat duplikat dalam dataset yang digunakan. Hal ini menunjukkan bahwa setiap entri data bersifat unik dan merepresentasikan observasi yang valid.

4.5. Cek Outlier



Gambar 4. Tren Inflasi Subkelompok Makanan Provinsi Sumatera Utara (2021–2025)

Outlier diidentifikasi menggunakan metode *Interquartile Range* (IQR) dan analisis visual melalui boxplot. Berdasarkan Gambar 4.3, sebagian besar komoditas menunjukkan adanya nilai ekstrem, terutama pada komoditas hortikultura seperti cabai merah, cabai rawit, dan bawang merah yang memiliki volatilitas harga tinggi. Sebaliknya, komoditas seperti beras, gula pasir, dan minyak goreng menunjukkan distribusi harga yang relatif stabil.

Meskipun terdapat nilai yang secara statistik teridentifikasi sebagai outlier, nilai tersebut tidak dihapus karena mencerminkan kondisi pasar yang aktual, seperti fluktuasi harga akibat faktor musiman dan gangguan pasokan. Oleh karena itu, seluruh data tetap dipertahankan agar model dapat menangkap dinamika harga secara lebih realistis.

4.6. Agregasi Data

Agregasi data dilakukan untuk menyelaraskan frekuensi data harga pangan harian dengan data inflasi bulanan. Harga harian diagregasi menjadi rata-rata bulanan pada tingkat provinsi, kemudian digabungkan dengan data inflasi berdasarkan periode waktu untuk membentuk dataset terpadu.

Selanjutnya, dihitung indeks harga per komoditas dengan menetapkan periode dasar, serta indeks harga pangan total berbobot sebagai representasi agregat. Indeks tersebut digunakan untuk menghitung inflasi pangan secara manual, yang menunjukkan pola sejalan dengan data inflasi resmi.

4.7. Rekayasa Fitur

Rekayasa fitur dilakukan untuk membentuk variabel prediktor yang relevan dalam pemodelan inflasi pangan. Variabel target ditetapkan sebagai inflasi pangan bulan berikutnya. Fitur yang digunakan meliputi lag inflasi (*inflasi_lag1*, *inflasi_lag3*), lag indeks harga pangan (*all_index_lag1*, *all_index_lag3*), serta perubahan indeks (*all_index_pctchg*).

Selain itu, dibentuk fitur perubahan harga (*percentage change*) untuk masing-masing komoditas pangan strategis guna menangkap pengaruh dinamika harga terhadap inflasi. Variabel musiman juga ditambahkan, yaitu bulan (*month*) dan indikator Idulfitri, untuk merepresentasikan pola musiman.

4.8. Pembagian Data

Selanjutnya, dilakukan pembagian data menggunakan pendekatan *time-series split*, di mana data tidak diacak tetapi disusun berdasarkan urutan waktu.

Tabel 3. Pembagian Data Training dan Testing

Kelompok Data	Jumlah Observasi	Jumlah Fitur
Training Set	41	17
Testing Set	11	17

Pendekatan ini bertujuan menjaga urutan temporal data serta menghindari *information leakage* dalam proses evaluasi model.

4.9. Pemodelan XGBoost

Tahap selanjutnya adalah membangun model prediksi inflasi pangan berdasarkan dataset hasil *feature engineering* pada tahap sebelumnya. Proses ini mencakup penerapan algoritma XGBoost sebagai model awal (*baseline*) serta optimasi model melalui *hyperparameter tuning* untuk meningkatkan kinerja prediksi.

Tabel 4. Parameter Model Baseline

n_estimators	300
learning_rate	0.05
max_depth	3
subsample	0.8
colsample_bytree	0.8

Pemodelan diawali dengan pembangunan model XGBoost baseline menggunakan parameter awal sebagai acuan untuk mengevaluasi performa awal model.

Tabel 5. Ruang Pencarian Hyperparameter XGBoost

Hyperparameter	Ruang Pencarian
n_estimators	{200, 300, 500}
learning_rate	{0.03, 0.05, 0.1}
max_depth	{2, 3, 4}
subsample	{0.7, 0.9}
colsample_bytree	{0.7, 0.9}

Selanjutnya, dilakukan proses *hyperparameter tuning* menggunakan metode GridSearchCV yang dikombinasikan dengan *TimeSeriesSplit* untuk menjaga urutan temporal data dan menghindari *information leakage*.

Tabel 6. Hyperparameter Terbaik

n_estimators	200
learning_rate	0.1
max_depth	4
subsample	0.9
colsample_bytree	0.9

Hasil dari proses tuning berupa kombinasi parameter terbaik yang menghasilkan model dengan performa optimal. Model tersebut kemudian digunakan pada tahap evaluasi untuk mengukur akurasi dan kemampuan generalisasi dalam memprediksi inflasi pangan.

4.10. Evaluasi Model

Evaluasi kinerja model dilakukan untuk menilai kemampuan XGBoost dalam memprediksi inflasi pangan pada data pengujian. Penilaian menggunakan tiga metrik utama, yaitu *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan *koefisien determinasi* (R^2).

Tabel 7. Perbandingan Kinerja Model XGBoost

Model	RMSE	MAE	R^2
XGBoost Baseline	0.5649	0.4176	0.8809
XGBoost Tuning Hyperparameter	0.5225	0.4530	0.8981

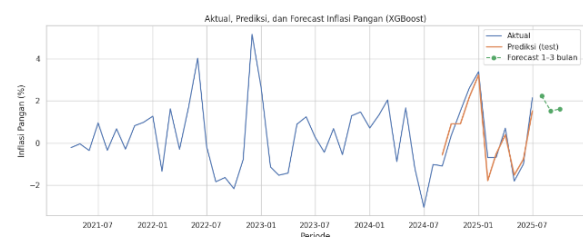
4.11. Peramalan (*Forecasting*) Inflasi Pangan

Tahap selanjutnya adalah peramalan inflasi pangan menggunakan model terbaik. Metode yang digunakan adalah *multi-step autoregressive forecasting*, di mana hasil prediksi pada satu periode digunakan sebagai input untuk periode berikutnya. Peramalan dilakukan untuk tiga periode ke depan.

Tabel 8. Hasil Peramalan Inflasi Pangan

Periode	Inflasi prediksi
2025-08-01	2.233745
2025-09-01	1.524878
2025-10-01	1.606632

Hasil peramalan menunjukkan bahwa inflasi pangan diprediksi mengalami penurunan dari 2,23% pada Agustus 2025 menjadi 1,52% pada September 2025, kemudian meningkat kembali menjadi 1,61% pada Oktober 2025. Pola ini menunjukkan bahwa inflasi pangan bersifat fluktuatif dengan kecenderungan melemah dalam jangka pendek.

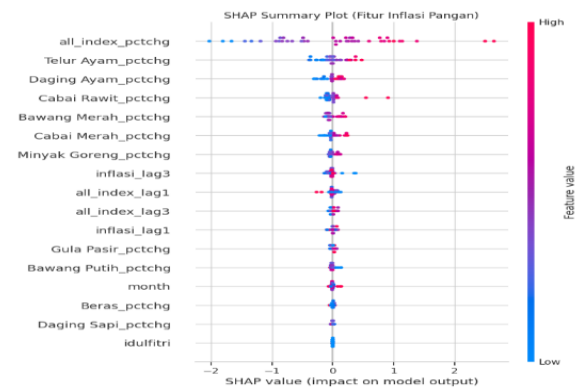


Gambar 5. Plot Forecast 3-periode ke depan

Kemudian, visualisasi hasil peramalan yang membandingkan data aktual, prediksi pada data uji, dan hasil forecast ditampilkan model XGBoost menunjukkan kemampuan yang cukup baik dalam mengikuti pola pergerakan data aktual, meskipun terdapat beberapa deviasi pada titik-titik tertentu, terutama pada periode dengan fluktuasi tinggi. Hasil forecast menunjukkan tren inflasi yang relatif stabil pada kisaran 1,5% hingga 2,2% dalam jangka pendek. Hal ini menunjukkan bahwa model memiliki kemampuan generalisasi yang baik dalam menangkap pola inflasi pangan, meskipun tetap memiliki keterbatasan dalam kondisi ekstrem.

4.12. Feature Importance

Untuk memahami kontribusi masing-masing variabel terhadap hasil prediksi, dilakukan analisis menggunakan metode SHAP (*Shapley Additive Explanations*).



Gambar 6. SHAP Summary

Hasil analisis menunjukkan bahwa variabel perubahan indeks harga pangan (*all_index_pctchg*) merupakan faktor yang paling dominan dalam memengaruhi inflasi pangan. Selain itu, perubahan harga komoditas seperti telur ayam, daging ayam, cabai rawit, dan bawang merah juga memberikan kontribusi signifikan. Sebaliknya, variabel lag inflasi serta beberapa komoditas seperti beras dan daging sapi memiliki pengaruh yang relatif lebih kecil. Temuan ini menunjukkan bahwa model lebih sensitif terhadap perubahan harga aktual dibandingkan pola historis, yang sejalan dengan karakteristik inflasi pangan sebagai komponen *volatile food*.

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma *Extreme Gradient Boosting* (XGBoost) mampu memodelkan dan memprediksi inflasi pangan di Provinsi Sumatera Utara dengan tingkat akurasi yang baik. Hasil evaluasi menunjukkan bahwa model XGBoost dengan optimasi hyperparameter memberikan kinerja yang lebih baik dibandingkan model baseline, dengan nilai RMSE sebesar 0,5225, MAE sebesar 0,4530, dan R^2 89,81% sebesar . Hal ini menunjukkan bahwa model memiliki kemampuan yang tinggi dalam menjelaskan variasi data inflasi pangan. Selain itu, analisis menggunakan metode SHAP menunjukkan bahwa variabel perubahan indeks harga pangan (*all_index_pctchg*) merupakan faktor yang paling dominan dalam memengaruhi inflasi, diikuti oleh perubahan harga komoditas strategis seperti telur ayam, daging ayam, cabai rawit, dan bawang merah. Untuk penelitian selanjutnya, disarankan untuk menambahkan variabel eksternal yang berpotensi memengaruhi inflasi pangan, seperti faktor cuaca, distribusi logistik, harga energi, dan kebijakan pemerintah, serta memperluas cakupan wilayah dan periode data erta membandingkan dengan metode lain.

DAFTAR PUSTAKA

- [1] M. Hafidz, M. Rizah Fahlevi, and S. Alfa Centauri, "Pengaruh Inflasi Terhadap Pertumbuhan Ekonomi: Perspektif Indonesia," *J. Keuang. Negara dan Kebijakan. Publik*, vol. 3, no. 1, pp. 17–26, 2023.
- [2] J. Cahyani, S. Mujahidin, and T. P. Fiqar, "Implementasi Metode Long Short Term Memory (LSTM) untuk Memprediksi Harga Bahan Pokok Nasional," *J. Sist. dan Teknol. Inf.*, vol. 11, no. 2, p. 346, 2023, doi: 10.26418/justin.v11i2.57395.
- [3] B. I. Ismaya and D. F. Anugrah, "Determinant Of Food Infaltion : The Case Of Indonesia," *Bull. Monet. Econ. Bank.*, vol. 21(1), pp. 81–94, 2018, doi: <https://doi.org/10.21098/bemp.v21i1.926>.
- [4] R. Jairin Abdullah, "Analisis Daya Beli Masyarakat terhadap Inflasi Barang dan Pertumbuhan Ekonomi Mikro," *Emanasi J. Ilmu Keislam. dan Sos.*, vol. 7, no. 2, pp. 27–40, 2024, [Online]. Available: <https://adpiks.or.id/ojs/index.php/emanasi/article/view/126>
- [5] F. Andriani and Roni Mohamad, "Urgensi Bank Sentral Indonesia Dalam Mengendalikan Laju Inflasi Melalui Sektor Riil," *J. Ekon. Syariah IAIN Sultan Amai Gorontalo*, vol. 3, no. 1, pp. 1–15, 2022, doi: 10.54045/mutawazin.v3i1.608.
- [6] DJPb Sumatera Utara, *Kajian Fiskal Regional Tahunan 2024*. Sumatera Utara, 2024. [Online]. Available: <https://ppa2.kanwildjpbsumut.com/kfr/2024/twiv/twiv2024.html>
- [7] BPS Sumatera Utara, "Inflasi Sumatera Utara: Desember 2024." [Online]. Available: <https://sumut.bps.go.id/id/pressrelease/2025/01/02/1309/desember-2024--inflasi-year-on-year--y-on-y--provinsi-sumatera-utara-sebesar-2-12-persen--inflasi-y-on-y-tertinggi-terjadi-di-kabupaten-labuhanbatu--yakni-sebesar-3-36-persen-.html>
- [8] Rahmanta, A. Sri Fajar, F. E. F., and S. R. S., "Pengaruh Fluktuasi Harga Komoditas Pangan Terhadap Inflasi di Provinsi Sumatera Utara," *Agrica (Jurnal Agribisnis Sumatera Utara)*, vol. 4, no. 1, pp. 238–249, 2020, doi: 10.47467/elmal.v4i1.1373.
- [9] P. D. Nindian, "Implementasi Holt-Winters Exponential Smoothing untuk Peramalan Harga Bahan Pangan di Kabupaten Pamekasan," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 11, no. 2, pp. 223–236, 2020, doi: 10.31849/digitalzone.v11i2.4797.
- [10] R. Azwina and M. Syahbudi, "Pengaruh Fluktuasi Harga Komoditas Pangan Terhadap Inflasi di Provinsi Sumatera Utara tahun (2019-2021)," *El-Mal J. Kaji. Ekon. Bisnis Islam*, vol. 4, no. 1, pp. 238–249, 2022, doi: 10.47467/elmal.v4i1.1373.
- [11] H. S. Afralia, T. Fahrudin, and Rochmawati,

- “Perbandingan Model Estimasi Inflasi Pangan Menggunakan Algoritma Naive Forecaster Dan Exponential Smoothing,” *J. Manaj. dan Akunt.*, vol. 18, no. 2, pp. 566–580, 2023, doi: 10.32534/jv.v18i2.4319.
- [12] F. M. Saragih and W. C. Wibowo, “Analysis of Food Security Index Predictions in Indonesia Using Machine Learning Approach,” *Agro Bali Agric. J.*, vol. 8, no. 2, pp. 377–393, 2025, [Online]. Available: <https://ejournal.unipas.ac.id/index.php/Agro/article/view/2302>
- [13] A. Gifty and D. Yang Li, “A Comparative Analysis of LSTM, ARIMA, XGBoost Algorithms in Predicting Stock Price Direction,” *Eng. Technol. J.*, vol. 09, no. 08, pp. 4978–4986, 2024, doi: 10.47191/etj/v9i08.50.
- [14] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-Augu, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.
- [15] R. R. Y. P. K. S. D. R. P. Y. D. S. M. D’Souza, “Time Series Forecasting for Energy Consumption Using XGBoost and LSTM,” *Asia Pacific Conf. Innov. Technol.*, pp. 1–8, 2024, doi: <https://doi.org/10.1109/APCIT62007.2024.10673558>.
- [16] BPS, “Statistik menurut Subjek.” [Online]. Available: <https://www.bps.go.id/id/statistics-table?subject=536&keyword=konsumen&page=1>
- [17] Bank Indonesia, “Inflasi.” [Online]. Available: <https://www.bi.go.id/id/fungsi-utama/moneter/inflasi/default.aspx>
- [18] World Bank, “Pembaruan Keamanan Pangan.” [Online]. Available: <https://www.worldbank.org/en/topic/agriculture/brief/food-security-update>
- [19] D. O. M. Abbas, “Forecasting with Machine Learning Methods,” *Adv. Stud. Theor. Appl. Econom.*, vol. 53, pp. 111–149, 2022, doi: 10.1007/978-3-031-15149-1_4.
- [20] M. Douglas C., C. L. Jennings, and M. Kulahci, “Introduction Time Series Analysis and Forecasting,” p. 671, 2015.
- [21] J. Friedman, “Greedy Function Approximation A Gradient Boosting Machine,” *Ann. Stat.*, pp. 1189–1232, 2001.
- [22] T. A. E. Putri, T. Widiyarih, and R. Santoso, “PENERAPAN TUNING HYPERPARAMETER RANDOMSEARCHCV PADA ADAPTIVE BOOSTING UNTUK PREDIKSI KELANGSUNGAN HIDUP PASIEN GAGAL JANTUNG Tita,” *J. GAUSSIAN*, vol. 11, pp. 397–406, 2023, doi: 10.14710/j.gauss.11.3.397-406.
- [23] M. R. Mubarak, M. Muliadi, and R. Herteno, “HYPER-PARAMETER TUNING PADA XGBOOST UNTUK PREDIKSI
- KEBERLANGSUNGAN HIDUP PASIEN GAGAL JANTUNG,” *KLIK (Kumpulan J. Ilmu Komputer)*, vol. 9, 2022, [Online]. Available: <https://garuda.kemdiktisaintek.go.id/documents/detail/2834085>
- [24] N. Amalia and Asmunin, “Optimasi Algoritma Random Forest dengan Hyperparameter Tuning Menggunakan GridSearchCV untuk Prediksi Nasabah Churn pada Industri Perbankan,” *J. Manaj. Inform.*, pp. 1–9, 2024, [Online]. Available: <https://ejournal.unesa.ac.id/index.php/jurnal-manajemen-informatika/article/view/60063>
- [25] C. Molnar, *Interpretable Machine Learning A Guide for Making Black Box Models Explainable*, 3rd ed. 2019. [Online]. Available: <https://christophm.github.io/interpretable-ml-book/>
- [26] S. M. Lundberg *et al.*, “From Local Explanations to Global Understanding with Explainable AI for Trees,” *Nat Mach Intell*, vol. 2, no. 1, pp. 56–67, 2020, doi: 10.1038/s42256-019-0138-9.From.
- [27] D. Chicco, M. J. Warrens, and G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,” *PeerJ Comput. Sci.*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.
- [28] S. I. Momo, M. Riajuliislam, and R. Hafiz, “Forecasting of Inflation Rate Contingent on Consumer Price Index: Machine Learning Approach,” *Lect. Notes Networks Syst.*, vol. 248, no. October, pp. 137–144, 2021, doi: 10.1007/978-981-16-3153-5_17.
- [29] M. Noorunnahar, A. H. Chowdhury, and F. A. Mila, “A tree based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh,” *PLoS One*, vol. 18, no. 3 March, pp. 1–15, 2023, doi: 10.1371/journal.pone.0283452.
- [30] M. Azkaenza, “Employing Machine Learning Techniques to Forecast Food Inflation in Indonesia: The Role of Food Price Indices,” 2023.
- [31] D. Riando and A. Afyati, “Implementation Of XGBoost Algorithm To Predict The Selling Price Of Cayenne Peppers In DKI Jakarta,” *J. Univers. Stud.*, vol. 4, no. 09, pp. 741–749, 2024.
- [32] A. M. Farrel, A. M. Rizki, and A. N. Sihananto, “Peramalan Tingkat Inflasi Di Indonesia Menggunakan Artificial Bee Colony Dan Xgboost,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4827.
- [33] F. Elfaladonna, I. G. T. Isa, D. Sartika, Yusniarti, and A. M. Putra, *Dasar Eksploratory Data Analisis (EDA)*. 2024. [Online]. Available: https://books.google.co.id/books/about/Buku_Ajar_Dasar_Eksploratory_Data_Analisi.html?id=n7ofEQAAQBAJ&redir_esc=y