

# Integrasi ARIMA, SARIMAX, XGBoost, dan Random Forest dalam Prediksi Harga Komoditas Pangan

## I. PENDAHULUAN

### A. Deskripsi Masalah

Fluktuasi harga pangan memiliki dampak yang luas terhadap ekonomi dan kebijakan publik. Sebagai contoh, lonjakan harga cabai merah di Indonesia sering kali menyebabkan inflasi pangan yang signifikan, mempengaruhi daya beli masyarakat, dan meningkatkan biaya produksi bagi pelaku industri makanan. Menurut data Badan Pusat Statistik (BPS), kenaikan harga pangan menyumbang lebih dari 30% terhadap inflasi keseluruhan di tahun-tahun tertentu. Oleh karena itu, diperlukan sistem prediksi harga yang akurat guna membantu perencanaan pasokan dan pengambilan kebijakan yang lebih efektif.

### B. Latar Belakang

Seiring dengan meningkatnya ketidakpastian dalam pasar pangan global, analisis faktor-faktor yang mempengaruhi fluktuasi harga bahan pangan menjadi semakin relevan. Perkembangan teknologi dalam bidang data science memungkinkan pemanfaatan dataset historis untuk membangun model prediktif yang dapat mengestimasi harga komoditas pangan dengan tingkat akurasi yang tinggi. Dataset yang digunakan dalam analisis ini mencakup berbagai variabel seperti tren harga historis, kondisi pasar, nilai tukar mata uang, serta indikator makroekonomi lainnya yang berperan dalam perubahan harga pangan di berbagai wilayah. Dengan penerapan teknik analisis data yang tepat, diharapkan dapat ditemukan pola-pola yang dapat digunakan untuk memahami dinamika pasar dan memperkirakan pergerakan harga di masa depan.

### C. Tujuan Analisis

Penelitian ini bertujuan untuk mengembangkan model prediksi harga komoditas pangan yang mampu memberikan estimasi harga dengan akurasi tinggi berdasarkan data historis yang tersedia. Selain itu, penelitian ini juga bertujuan untuk:

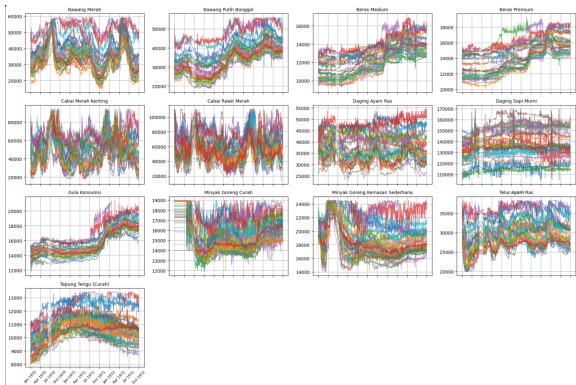
1. Mengidentifikasi faktor-faktor utama yang mempengaruhi volatilitas harga komoditas pangan.
2. Menguji efektivitas berbagai teknik data science dalam membangun model prediksi harga.
3. Memberikan wawasan yang lebih dalam mengenai pola pergerakan harga pangan guna mendukung pengambilan keputusan berbasis data.
4. Berkontribusi dalam upaya meningkatkan ketahanan pangan dan stabilitas ekonomi melalui pemanfaatan prediksi harga yang lebih akurat.

Dengan demikian, penelitian ini diharapkan dapat menjadi landasan dalam perencanaan strategi pengelolaan rantai pasokan pangan dan mitigasi risiko volatilitas harga, serta memberikan manfaat bagi berbagai pihak yang berkepentingan dalam sektor pangan.

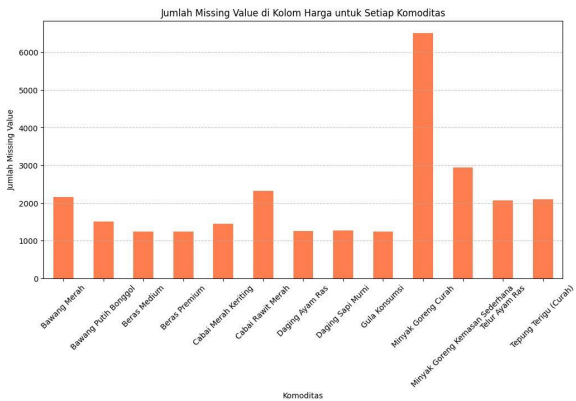
## II. METODE ANALISIS

### A. Understanding Data

GRAFIK 2.1 GRAFIK TIME PLOT PER KOMODITAS



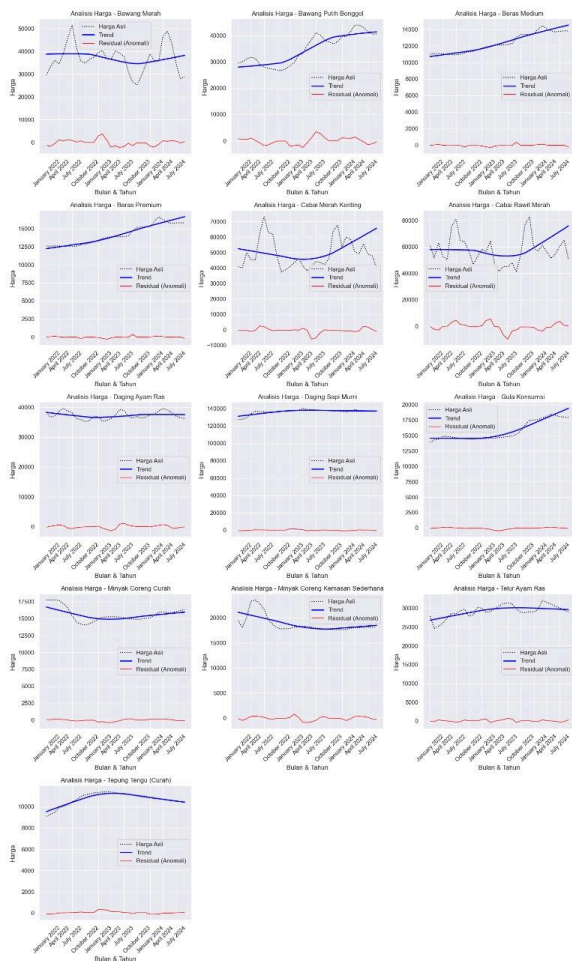
TABEL 1.1 TABEL MISSING VALUE PER KOMODITAS



Dataset yang digunakan dalam penelitian ini merupakan data time series yang mencatat harga berbagai komoditas pangan dalam rentang waktu tertentu. Data time series memiliki karakteristik unik, di mana pola tren dan musiman dapat diamati dalam pergerakan harga. Oleh karena itu, analisis eksploratif terhadap data sangat diperlukan sebelum melakukan pemodelan prediksi.

Berdasarkan hasil eksplorasi data, ditemukan bahwa beberapa komoditas memiliki jumlah missing value yang signifikan dalam kolom harga. Grafik pertama menunjukkan distribusi missing value untuk setiap komoditas, di mana minyak goreng curah memiliki jumlah missing value tertinggi dibandingkan komoditas lainnya, diikuti oleh minyak goreng kemasan sederhana. Selain itu, beberapa komoditas seperti bawang merah, cabai rawit merah, dan beras juga memiliki jumlah missing value yang cukup besar. Keberadaan missing value dalam jumlah besar dapat mempengaruhi performa model prediksi jika tidak ditangani dengan baik.

GRAFIK 2.2 TREK HARGA PER KOMODITAS



Berdasarkan hasil eksplorasi data, sebagian besar komoditas mengalami kenaikan harga dalam jangka panjang, seperti beras premium, gula konsumsi, dan daging sapi murni. Hal ini menunjukkan adanya inflasi atau peningkatan permintaan terhadap komoditas tersebut. Sebaliknya, beberapa komoditas mengalami penurunan atau stabil dalam tren harga, seperti minyak goreng curah dan minyak goreng kemasan sederhana. Fenomena ini dapat disebabkan oleh intervensi pemerintah atau perubahan suplai pasar.

Selain itu, harga cabai merah keriting dan cabai rawit merah menunjukkan fluktuasi yang cukup tinggi, tetapi secara jangka panjang tetap mengalami peningkatan. Faktor-faktor seperti musim panen dan kondisi cuaca dapat menjadi penyebab utama dari volatilitas harga tersebut. Selain itu, ditemukan bahwa beberapa komoditas menunjukkan residual atau anomali harga yang cukup besar, seperti cabai merah dan minyak goreng. Ini menandakan adanya lonjakan atau penurunan harga yang tajam pada periode tertentu. Sementara itu, komoditas seperti beras premium dan gula konsumsi memiliki residual yang lebih kecil, menunjukkan bahwa harga mereka cenderung lebih stabil tanpa banyak fluktuasi mendadak.

Beberapa komoditas menunjukkan pola musiman yang jelas, di mana harga mengalami fluktuasi yang berulang dalam periode tertentu. Misalnya, cabai merah dan minyak goreng memiliki pola harga yang berulang, yang kemungkinan besar disebabkan oleh faktor cuaca, siklus panen, atau permintaan musiman seperti menjelang hari raya. Sebaliknya, beras dan daging ayam ras menunjukkan pergerakan harga yang lebih stabil, tanpa pola musiman yang signifikan.

## B. Pengimputan Missing Values

Dalam menangani missing value pada data time series, terdapat berbagai metode yang dapat digunakan. Dari berbagai metode yang tersedia, interpolasi linier merupakan pilihan yang paling ideal. Metode ini memiliki beberapa keunggulan utama yang membuatnya lebih efektif

dibandingkan metode lainnya.

Pertama, interpolasi linier mampu mempertahankan pola *trend* dan musiman yang ada dalam data. Dalam konteks harga pangan, pola ini sangat penting karena harga seringkali dipengaruhi oleh tren inflasi dan faktor musiman seperti panen dan hari besar keagamaan. Dengan menggunakan interpolasi, nilai yang hilang dapat diisi dengan mempertimbangkan kecenderungan alami dari data, sehingga pola historis tetap terjaga.

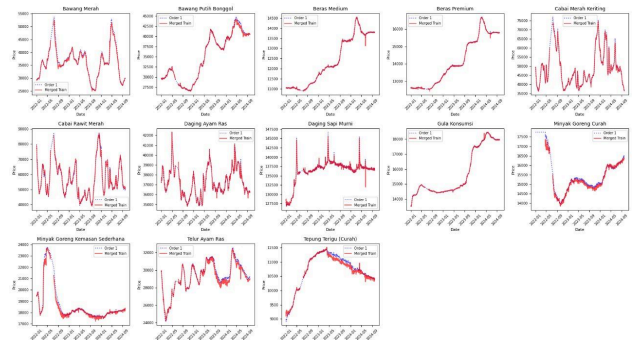
Kedua, interpolasi menghindari distorsi data yang dapat terjadi pada metode lain seperti moving average atau *KNN imputation*. Beberapa metode tersebut cenderung mengubah pola harga secara tidak alami, yang dapat menyebabkan perubahan distribusi data. Sebaliknya, interpolasi hanya menggunakan titik data yang tersedia untuk mengisi nilai yang hilang, tanpa mengubah pola aslinya.

Selain itu, interpolasi tidak memerlukan data eksternal atau model yang kompleks seperti yang dibutuhkan dalam metode *KNN imputation* atau *time series forecasting* (misalnya ARIMA dan *SARIMA*). Hal ini menjadikannya metode yang lebih efisien dan praktis, terutama dalam kasus di mana data tambahan sulit diperoleh atau model prediksi yang lebih kompleks tidak diperlukan.

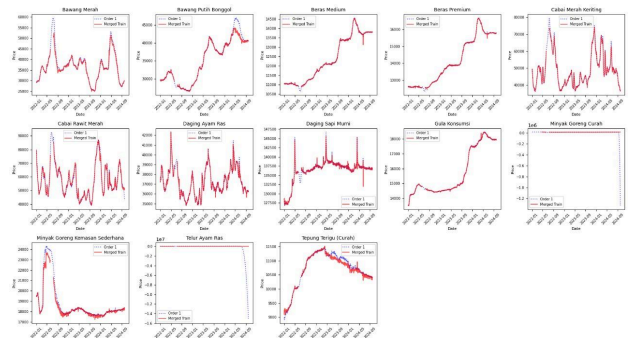
Terakhir, interpolasi linier memiliki akurasi tinggi dalam menangani missing value dalam jumlah kecil hingga sedang yang tersebar secara sporadis. Jika missing value tidak terlalu banyak, interpolasi dapat memberikan estimasi yang sangat mendekati nilai aslinya dibandingkan metode lain.

Setelah metode interpolasi diterapkan, evaluasi terhadap hasil imputasi perlu dilakukan. Salah satu cara evaluasi adalah dengan melakukan visualisasi ulang untuk memastikan bahwa pola harga tetap masuk akal setelah dilakukan imputasi. Dengan pendekatan ini, data yang digunakan dalam penelitian dapat lebih bersih dan siap untuk tahap pemodelan prediksi harga komoditas pangan.

GRAFIK 2.3 HASIL INTERPOLASI MENGGUNAKAN ORDE 1



GRAFIK 2.4 HASIL INTERPOLASI MENGGUNAKAN ORDE 3



Interpolasi polinomial orde 3 (kubik) cenderung menghasilkan kurva yang sangat fluktuatif dan kurang mampu melakukan generalisasi, terutama pada data yang memiliki tren lebih stabil. Hal ini terlihat pada minyak goreng curah, di mana interpolasi orde 3 kurang mampu mengisi data dengan baik karena model ini lebih cocok untuk tren yang fluktuatif. Sifat polinomial kubik yang lebih sensitif terhadap variasi data dapat menyebabkan overfitting, sehingga pola yang dihasilkan terlalu responsif terhadap perubahan kecil dalam data. Akibatnya, interpolasi kubik tidak selalu menangkap tren umum dengan baik, berbeda dengan metode orde lebih rendah seperti interpolasi linier yang lebih sederhana dan cenderung lebih stabil dalam menangkap pola data yang lebih halus.

Hasil imputasi menggunakan interpolasi orde 1 menunjukkan pola yang lebih logis dan realistis dalam merepresentasikan data. Metode ini bekerja dengan menghubungkan titik-titik data secara linier, sehingga mampu mengisi nilai yang hilang dengan pendekatan yang lebih stabil dan tidak berlebihan dalam menangkap fluktuasi kecil. Dengan demikian, interpolasi orde 1 lebih sesuai untuk tren harga yang cenderung bergerak secara gradual, memastikan hasil yang lebih akurat dan dapat diandalkan untuk analisis lebih lanjut.

### C. Pemilihan Model

Dalam penelitian ini, dipilih empat model utama untuk melakukan prediksi harga komoditas pangan. Pemilihan model ini dilakukan berdasarkan kemampuan masing-masing dalam menangani karakteristik data time series yang memiliki tren, musiman, serta hubungan non-linear.

#### 1. ARIMA (AutoRegressive Integrated Moving Average)

Model ARIMA dipilih karena mampu menangani data yang memiliki tren sederhana dan tidak terlalu dipengaruhi oleh faktor musiman yang kompleks. Model ini efektif dalam menganalisis pola linier dalam data time series.

#### 2. SARIMAX (Seasonal AutoRegressive Integrated Moving Average with Exogenous Variables)

SARIMAX merupakan pengembangan dari ARIMA yang dapat menangani data dengan pola musiman serta mempertimbangkan faktor eksternal yang dapat mempengaruhi harga komoditas, seperti cuaca atau kebijakan pemerintah.

#### 3. XGBoost (Extreme Gradient Boosting)

Model XGBoost digunakan karena memiliki kemampuan untuk menangani pola non-linear yang lebih kompleks. Algoritma ini dikenal dengan kemampuannya dalam menangani data besar serta menghasilkan prediksi yang akurat melalui teknik gradient boosting.

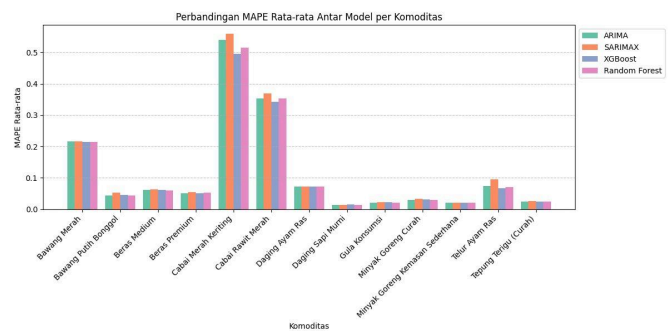
#### 4. Random Forest

Random Forest dipilih karena mampu menangani hubungan non-linear dalam data, serta lebih toleran terhadap outlier. Model ini juga dapat memanfaatkan fitur tambahan yang mungkin berpengaruh terhadap harga komoditas.

Dengan membandingkan performa keempat model ini, dapat diketahui metode mana yang paling efektif dalam memprediksi harga komoditas pangan berdasarkan pola historis yang telah diamati. Evaluasi dilakukan menggunakan metrik error seperti RMSE (Root Mean Squared Error) dan MAE (Mean Absolute Error) untuk menentukan model dengan performa terbaik. Hasil dari analisis ini akan menjadi dasar dalam pemilihan model akhir yang digunakan dalam prediksi harga pangan.

## III. HASIL DAN PEMBAHASAN

TABEL 3.1 PERBANDINGAN MAPE PER MODEL



Evaluasi model dilakukan dengan menggunakan Mean Absolute Percentage Error (MAPE) untuk mengukur tingkat kesalahan prediksi harga komoditas. Hasil MAPE menunjukkan variasi performa antar model, tergantung pada karakteristik data tiap komoditas. Secara umum, model yang digunakan, yaitu ARIMA, SARIMAX, XGBoost, dan Random Forest, memiliki keunggulan masing-masing dalam menangani berbagai jenis pola data.

ARIMA menunjukkan performa yang baik pada komoditas dengan pola tren sederhana, seperti beras dan minyak goreng. Namun, model ini cenderung memiliki error lebih besar pada komoditas dengan pola musiman yang kuat, seperti cabai merah keriting dan cabai rawit merah. SARIMAX, yang mempertimbangkan faktor musiman, menunjukkan perbaikan performa dibandingkan ARIMA pada beberapa komoditas dengan pola musiman seperti cabai dan daging ayam ras. Namun, SARIMAX masih memiliki keterbatasan dalam menangani pola harga yang non-linear.

XGBoost memberikan performa lebih baik pada komoditas dengan pola harga yang lebih kompleks dan non-linear, seperti bawang merah dan tepung terigu. Namun, model ini masih mengalami kesulitan dalam menangkap pola musiman yang tidak stabil. Random Forest memberikan performa yang cukup kompetitif dalam menangani komoditas dengan pola harga yang lebih acak dan dipengaruhi oleh banyak faktor eksternal. Meski begitu, model ini dapat mengalami overfitting jika jumlah fitur tidak dikelola dengan baik.

Komoditas dengan pola *trend* stabil lebih cocok dengan pendekatan model statistik seperti ARIMA dan SARIMAX karena pola kenaikan atau penurunan harga dapat diprediksi dengan cukup baik menggunakan metode berbasis regresi waktu. Komoditas dengan pola harga fluktuatif lebih sulit diprediksi dengan ARIMA atau SARIMAX karena kedua model ini mengasumsikan pola yang lebih stabil. Model berbasis machine learning seperti XGBoost dan Random Forest lebih fleksibel dalam menangkap pola harga yang tidak teratur. Komoditas dengan faktor eksternal yang dominan, seperti kebijakan pemerintah atau cuaca, sulit diprediksi oleh model yang hanya mengandalkan data historis. Oleh karena itu, model yang dapat mengakomodasi variabel eksternal, seperti SARIMAX, cenderung memiliki kinerja lebih baik dibandingkan ARIMA murni.

Untuk meningkatkan akurasi prediksi harga komoditas, terdapat beberapa langkah yang dapat diterapkan dalam proses pemodelan. Salah satu langkah yang penting adalah optimasi hyperparameter, terutama untuk model berbasis machine learning seperti XGBoost dan Random Forest. Model-model ini sangat bergantung pada parameter yang dipilih, sehingga tuning hyperparameter yang lebih optimal dapat membantu menghindari overfitting serta meningkatkan generalisasi terhadap data baru. Selain itu,



penambahan variabel eksternal juga dapat menjadi strategi yang efektif. Faktor-faktor seperti curah hujan, tingkat inflasi, atau kebijakan impor memiliki pengaruh signifikan terhadap harga komoditas dan dapat dimasukkan sebagai variabel tambahan dalam model SARIMAX maupun model machine learning untuk meningkatkan akurasi prediksi.

Lebih lanjut, penerapan model hybrid menjadi pendekatan lain yang patut dipertimbangkan. Kombinasi model statistik dan machine learning memungkinkan pemanfaatan keunggulan dari masing-masing metode, misalnya dengan menggunakan ARIMA untuk menangkap tren jangka panjang yang lebih stabil, sementara XGBoost digunakan untuk mengidentifikasi pola jangka pendek yang lebih kompleks dan dinamis. Selain itu, dalam mengevaluasi performa model, penggunaan metrik tambahan seperti Root Mean Squared Error (RMSE) atau Symmetric Mean Absolute Percentage Error (SMAPE) dapat memberikan gambaran yang lebih komprehensif terhadap cara model menangani error pada skala yang berbeda. Dengan memahami kelebihan dan kekurangan dari setiap model serta menerapkan perbaikan yang sesuai, analisis prediksi harga komoditas dapat semakin akurat dan memberikan wawasan yang lebih mendalam bagi pengambil kebijakan serta pelaku industri pangan.

Implikasi dari temuan ini berkaitan erat dengan tujuan awal penelitian, yaitu mengembangkan model yang dapat memprediksi harga komoditas pangan dengan tingkat akurasi yang baik untuk membantu pengambilan keputusan. Dengan memahami kekuatan dan keterbatasan masing-masing model, pemilihan model prediksi dapat disesuaikan berdasarkan karakteristik komoditas yang dianalisis. Untuk komoditas dengan volatilitas tinggi, pendekatan hybrid yang menggabungkan metode statistik dan machine learning dapat menjadi solusi yang lebih efektif. Sementara itu, untuk komoditas dengan tren harga yang lebih stabil, model statistik seperti ARIMA atau SARIMAX dapat memberikan hasil yang cukup baik dengan computational cost yang lebih rendah.

Selain itu, hasil penelitian ini juga memberikan wawasan bagi pemangku kebijakan dan pelaku industri pangan dalam menyusun strategi stabilisasi harga dan perencanaan pasokan. Dengan model prediksi yang lebih akurat, intervensi kebijakan seperti impor atau subsidi dapat direncanakan dengan lebih tepat guna untuk mengantisipasi lonjakan harga yang tidak terduga. Implikasi ini menegaskan pentingnya pengembangan model prediktif yang tidak hanya mengandalkan satu pendekatan, tetapi juga mempertimbangkan faktor eksternal dan karakteristik unik dari setiap komoditas untuk mencapai hasil yang lebih optimal.

#### IV. KESIMPULAN

Studi ini mengonfirmasi bahwa pemilihan model prediksi harus disesuaikan dengan karakteristik komoditas yang dianalisis. Model statistik seperti ARIMA dan SARIMAX cocok untuk tren harga yang lebih stabil, sementara model machine learning lebih efektif dalam menangani volatilitas harga yang tinggi. Sebagai rekomendasi, implementasi model dalam kebijakan pangan dapat dilakukan dengan mengembangkan sistem peringatan dini (early warning system) bagi pemerintah untuk mendeteksi potensi lonjakan harga. Selain itu, sektor bisnis seperti ritel dan manufaktur pangan dapat memanfaatkan model ini untuk perencanaan stok dan strategi harga yang lebih adaptif terhadap dinamika pasar. Dengan pendekatan yang lebih cermat dalam pemodelan data dan pemilihan metode yang sesuai, prediksi harga komoditas dapat menjadi alat yang lebih andal dalam mendukung stabilitas ekonomi dan ketahanan pangan nasional.