

Perbandingan Kinerja Model Forecasting Nilai Perdagangan Komoditas HS pada Evaluasi Time-Based

Ridwan Dwi Irawan^{1*}, Marta Ardiyanto², Ringgo Ismoyo Buwono³, Faulinda Ely Nastiti⁴

¹ Fakultas Ilmu Komputer, Teknologi Rekayasa Perangkat Lunak, Universitas Duta Bangsa Surakarta, Surakarta, Indonesia

² Fakultas Ilmu Komputer, Sistem Informasi, Universitas Duta Bangsa Surakarta, Surakarta, Indonesia

³ Fakultas Ilmu Komputer, Teknik Industri, Universitas Duta Bangsa Surakarta, Surakarta, Indonesia

⁴ Fakultas Ilmu Komputer, Teknologi Rekayasa Perangkat Lunak, Universitas Duta Bangsa Surakarta, Surakarta, Indonesia

Email: ^{1,*}ridwan_dwiirawan@udb.ac.id, ²marta.ardiyanto@udb.ac.id, ³ringgo_ismoyo@udb.ac.id, ⁴faulinda@udb.ac.id

Email Penulis Korespondensi: ridwan_dwiirawan@udb.ac.id

Submitted 28-05-2026; Accepted 29-06-2026; Published 30-06-2026

Abstrak

Ketidakpastian ekonomi global, perubahan rezim perdagangan, dan gangguan rantai pasok membuat forecasting nilai perdagangan ekspor-impor semakin kompleks. Penelitian ini membandingkan kinerja Random Forest, Extra Trees Regression, dan SARIMA dalam memprediksi nilai perdagangan bulanan komoditas HS sektor pakaian dan alas kaki, yaitu HS 61, HS 62, HS 63, dan HS 64. Data penelitian bersumber dari Badan Pusat Statistik (BPS) dan disusun sebagai deret waktu bulanan dengan skema leakage-safe melalui pembagian train-validation-test berbasis waktu. Pemodelan dilakukan menggunakan 19 fitur prediktor yang mencakup fitur historis, statistik lokal, kalender-musiman, dan variabel eksogen, kemudian setiap model dituning pada validation set sebelum diuji pada holdout test. Evaluasi dilakukan menggunakan MAE, RMSE, MAPE, sMAPE, dan wMAPE dengan MAPE sebagai metrik utama. Kontribusi utama penelitian ini terletak pada komparasi tiga model forecasting dalam protokol evaluasi time-based yang setara dan replikabel pada data perdagangan komoditas HS, sehingga hasil seleksi model lebih representatif untuk kondisi implementasi nyata. Hasil penelitian menunjukkan bahwa Random Forest memperoleh MAPE terbaik sebesar 22,7746%, sedikit lebih baik dibanding Extra Trees Regression sebesar 22,9469%, sedangkan SARIMA menghasilkan 28,9794%. Temuan ini menunjukkan bahwa model ensemble tree lebih adaptif terhadap data perdagangan yang fluktuatif, sementara SARIMA tetap relevan sebagai baseline statistik untuk pola musiman yang terstruktur.

Kata Kunci: Forecasting; Time Series; Evaluasi Time-Based; Industri; SARIMA; Komoditas HS.

Abstract

Global economic uncertainty, trade-regime shifts, and supply-chain disruptions have made export-import trade-value forecasting increasingly complex. This study compares the performance of Random Forest, Extra Trees Regression, and SARIMA in predicting monthly trade values of HS commodities in the apparel and footwear sector, covering HS 61, HS 62, HS 63, and HS 64. The dataset was obtained from Indonesia's Central Bureau of Statistics (BPS) and organized as a monthly time series using a leakage-safe workflow through a time-based train-validation-test split. The modeling stage employed 19 predictive features consisting of historical, local statistical, calendar-seasonal, and exogenous variables, and each model was tuned on the validation set before being evaluated on the holdout test. Performance was assessed using MAE, RMSE, MAPE, sMAPE, and wMAPE, with MAPE as the primary ranking metric. The main contribution of this study lies in providing a fair and replicable comparison of three forecasting models under a time-based evaluation protocol for HS commodity trade data, making the model selection results more representative of real implementation settings. The results show that Random Forest achieved the best MAPE at 22.7746%, slightly outperforming Extra Trees Regression at 22.9469%, while SARIMA recorded 28.9794%. These findings indicate that tree-based ensemble models are more adaptive to volatile trade data, whereas SARIMA remains relevant as a statistical baseline for structured seasonal patterns.

Keywords: Forecasting; Time Series; Time-Based Evaluation; Industry; SARIMA; HS Commodities.

1. PENDAHULUAN

Pemilihan model *forecasting* yang tepat menjadi persoalan utama ketika deret waktu menunjukkan volatilitas, perubahan level, pengaruh faktor eksternal, dan kemungkinan perubahan rezim dari waktu ke waktu. Dalam situasi seperti ini, model yang tampak baik pada satu dataset belum tentu mempertahankan performa yang sama ketika diuji dengan protokol evaluasi yang berbeda.

Urgensi penelitian ini terletak pada karakter data perdagangan itu sendiri. Data nilai ekspor-impor bulanan cenderung tidak sepenuhnya stabil karena dipengaruhi kurs, harga global, biaya logistik, kebijakan perdagangan, dan shock eksternal [1]-[6]. Pada seri seperti ini, error prediksi biasanya meningkat ketika terjadi lonjakan tajam, perubahan musiman, atau pergeseran hubungan antara target dan variabel penjelas.

Objek penelitian menggunakan komoditas HS, yaitu *Harmonized System* yang merupakan standar klasifikasi barang dalam perdagangan internasional. Empat kelompok yang dianalisis adalah HS 61 (*articles of apparel and clothing accessories, knitted or crocheted*), HS 62 (*articles of apparel and clothing accessories, not knitted or crocheted*), HS 63 (*other made-up textile articles*), dan HS 64 (*footwear, gaiters, and the like*). Keempat kelompok ini dipilih karena mewakili sektor pakaian dan alas kaki yang aktif dalam arus perdagangan dan memiliki pola bulanan yang cukup dinamis.

Model statistik klasik seperti SARIMA tetap penting karena kuat untuk pola musiman yang relatif terstruktur. Namun, pada data yang lebih nonlinear dan dipengaruhi banyak faktor eksternal, model ensemble tree seperti *Random Forest* dan *Extra Trees Regression* sering lebih fleksibel dalam menangkap interaksi antarfitur tanpa asumsi bentuk fungsi yang ketat.

Berdasarkan kebutuhan tersebut, penelitian ini bertujuan membandingkan *Random Forest*, *Extra Trees Regression*, dan SARIMA untuk menentukan model dengan error relatif paling rendah dan performa paling konsisten pada evaluasi berbasis waktu. Kontribusi utama penelitian terletak pada komparasi model yang fair, penggunaan multi-metrik, serta penjelasan yang lebih eksplisit tentang kompleksitas data dan prosedur validasi. Agar hasil dapat dipertanggungjawabkan, seluruh proses pembentukan fitur, pelatihan, tuning, dan pengujian dipisahkan secara ketat berdasarkan urutan waktu. Melalui desain *time-based* yang *leakage-safe*, penelitian ini berupaya menghasilkan dasar seleksi model yang lebih realistis untuk implementasi forecasting pada data perdagangan komoditas HS.

Penelitian *forecasting* beberapa tahun terakhir menunjukkan bahwa perbandingan model perlu diletakkan pada karakter data dan desain evaluasi, bukan pada popularitas algoritma semata [4], [7]. Tinjauan komprehensif mutakhir menegaskan bahwa performa model deret waktu sangat sensitif terhadap pola musiman, noise, nonlinieritas, dan strategi validasi yang digunakan [4].

Pada konteks ekonomi dan perdagangan, dinamika eksternal seperti geopolitik, harga global, dan gangguan logistik dapat mengubah bentuk seri secara tiba-tiba [1]-[3], [6]. Karena itu, data perdagangan merupakan konteks yang relevan untuk menguji apakah model mampu mempertahankan akurasi pada seri yang tidak sepenuhnya stabil.

Secara metodologis, SARIMA tetap menjadi *baseline* penting untuk data dengan pola tren-musiman yang jelas, sedangkan *Random Forest* dan *Extra Trees Regression* banyak dipilih ketika hubungan antarvariabel bersifat nonlinier dan melibatkan banyak fitur turunan waktu [4], [7]-[9]. Pendekatan komparatif antar-model dalam protokol yang identik karenanya lebih kuat dibanding menetapkan model terbaik sejak awal berdasarkan asumsi teoritis saja. Sejumlah studi juga menunjukkan bahwa kualitas eksperimen forecasting tidak cukup dinilai dari skor error akhir, tetapi juga dari ketepatan desain evaluasinya. Jika split data, pembentukan fitur, atau tuning tidak mengikuti urutan waktu, maka estimasi performa dapat menjadi terlalu optimistis dan tidak mewakili kondisi implementasi nyata [8]-[11].

Selain desain evaluasi, karakteristik data dan pemilihan metrik juga menentukan interpretasi hasil. MAPE, sMAPE, dan wMAPE berguna ketika penelitian menekankan error relatif, sedangkan MAE dan RMSE penting untuk membaca deviasi pada skala absolut. Oleh karena itu, keputusan pemilihan model sebaiknya didasarkan pada konsistensi lintas metrik, bukan pada satu ukuran tunggal [11]-[14]. Selain literatur *forecasting* dan evaluasi deret waktu, daftar pustaka penelitian ini juga memuat sejumlah studi Explainable AI yang relevan sebagai landasan konseptual, terutama untuk menegaskan arah pengembangan model prediksi yang tidak hanya akurat, tetapi juga transparan [6], [15]-[26], [36]-[37]. Walaupun eksperimen pada artikel ini masih berfokus pada komparasi akurasi dan validitas evaluasi *time-based*, keberadaan referensi tersebut penting untuk menunjukkan posisi penelitian dalam jalur pengembangan forecasting yang menuju model lebih interpretable dan dapat dipertanggungjawabkan.

Gap penelitian masih terlihat pada studi yang secara eksplisit membandingkan model statistik klasik dan ensemble tree pada data perdagangan komoditas HS dalam satu protokol evaluasi *time-based* yang *leakage-safe*. Padahal, jenis perbandingan ini penting untuk menjelaskan apakah keunggulan model tertentu benar-benar berasal dari karakter model atau hanya akibat perlakuan eksperimen yang tidak setara.

Tabel 1. Perbandingan Penelitian Terdahulu dan Posisi Studi Ini

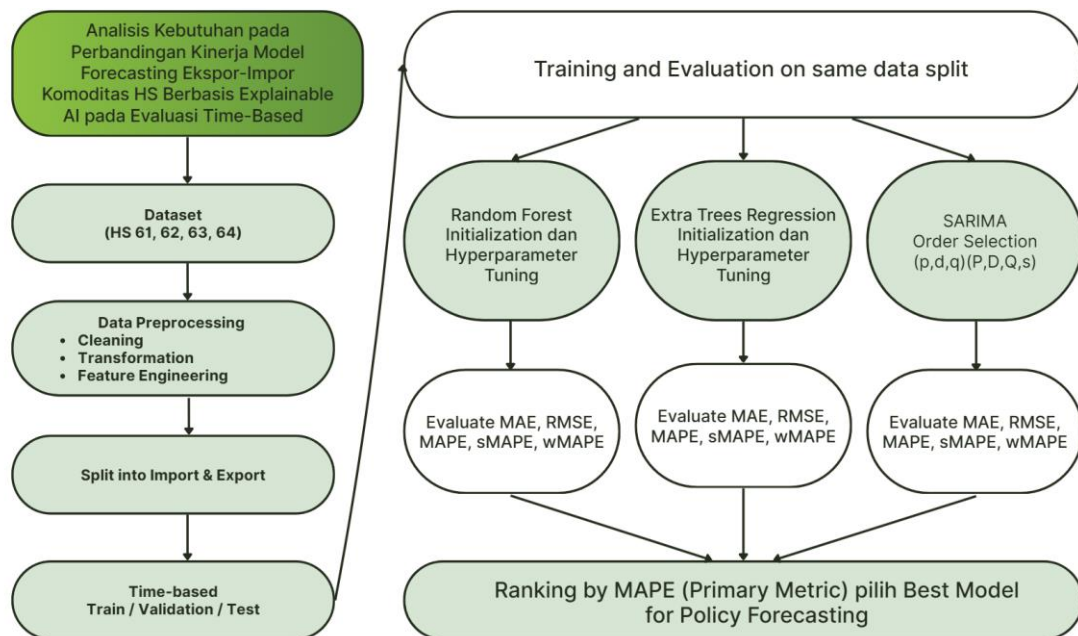
Peneliti	Objek Data	Metode	Skema Evaluasi	Temuan utama	Gap terhadap studi ini
Rincon-Yanez et al. [1]	International trade flows	Knowledge graph embeddings	Prediksi trade flow lintas negara	Akurat untuk relasi perdagangan kompleks	Tidak membandingkan model statistik klasik dan ensemble tree pada komoditas HS bulanan
Sellami et al. [5]	International trade flows	Graph Neural Networks	Eksperimen prediksi berbasis graph	Menunjukkan potensi model modern pada data perdagangan	Belum menekankan evaluasi leakage-safe berbasis time split untuk komoditas HS
Celik et al. [215]	Harga komoditas pertanian	Hybrid statistical-learning dan interpretable ML	Komparasi model pada data komoditas	ML cenderung adaptif pada data fluktuatif	Bukan pada data ekspor-impor HS dan tidak fokus pada fairness protokol evaluasi time-based
Studi ini	Nilai perdagangan komoditas HS 61-64	Random Forest, Extra Trees Regression, SARIMA	Train-validation-test berbasis waktu yang leakage-safe	Menyediakan komparasi fair dan multi-metrik pada data perdagangan bulanan	Menekankan seleksi model yang replikabel dan relevan untuk implementasi forecasting perdagangan

Dibandingkan penelitian terdahulu yang umumnya berfokus pada satu keluarga model, satu jenis metrik, atau evaluasi yang belum sepenuhnya menempatkan urutan waktu sebagai batas eksperimen, penelitian ini menempatkan Random Forest, Extra Trees Regression, dan SARIMA pada protokol *time-based* yang identik, leakage-safe, dan berbasis multi-metrik. Dengan demikian, kontribusi studi ini tidak hanya terletak pada pemilihan model dengan error relatif terbaik, tetapi juga pada penyediaan dasar seleksi model yang lebih fair, replikabel, dan relevan untuk implementasi forecasting perdagangan komoditas HS.

2. METODOLOGI PENELITIAN

2.1 Desain Penelitian

Penelitian ini menggunakan desain kuantitatif komparatif berbasis eksperimen forecasting time series. Tujuan desain ini adalah membandingkan performa Random Forest, Extra Trees Regression, dan SARIMA pada dataset, horizon prediksi, fitur, dan metrik evaluasi yang identik, sehingga perbedaan hasil yang muncul dapat diatribusikan pada karakter model, bukan pada perbedaan perlakuan eksperimen. Desain semacam ini penting dalam studi deret waktu karena validitas komparasi sangat dipengaruhi oleh kesetaraan protokol evaluasi dan konsistensi urutan waktu yang digunakan [4], [7], [8].



Gambar 1. Tahapan Metodologi

Pada Gambar 1 ditunjukkan alur eksperimen komparatif tiga metode peramalan yang diterapkan secara konsisten pada dataset yang sama. Proses dimulai dari penyusunan seri bulanan komoditas HS 61, HS 62, HS 63, dan HS 64, dilanjutkan dengan pembersihan data, rekayasa fitur temporal, pemisahan train-validation-test berbasis waktu, pelatihan model, tuning parameter, dan evaluasi pada holdout test.

Setup eksperimen menempatkan seluruh model pada skenario yang setara. *Random Forest*, *Extra Trees Regression*, dan SARIMA dijalankan pada split yang identik; model *ensemble* dituning pada *validation set* melalui parameter inti seperti jumlah pohon, kedalaman pohon, dan ukuran minimum sampel, sedangkan SARIMA dipilih melalui pencarian orde non-musiman dan musiman yang paling sesuai pada *validation set*. Seluruh model kemudian dievaluasi menggunakan MAE, RMSE, MAPE, sMAPE, dan wMAPE, dengan MAPE sebagai dasar utama pemeringkatan karena penelitian ini menekankan kestabilan *error* relatif.

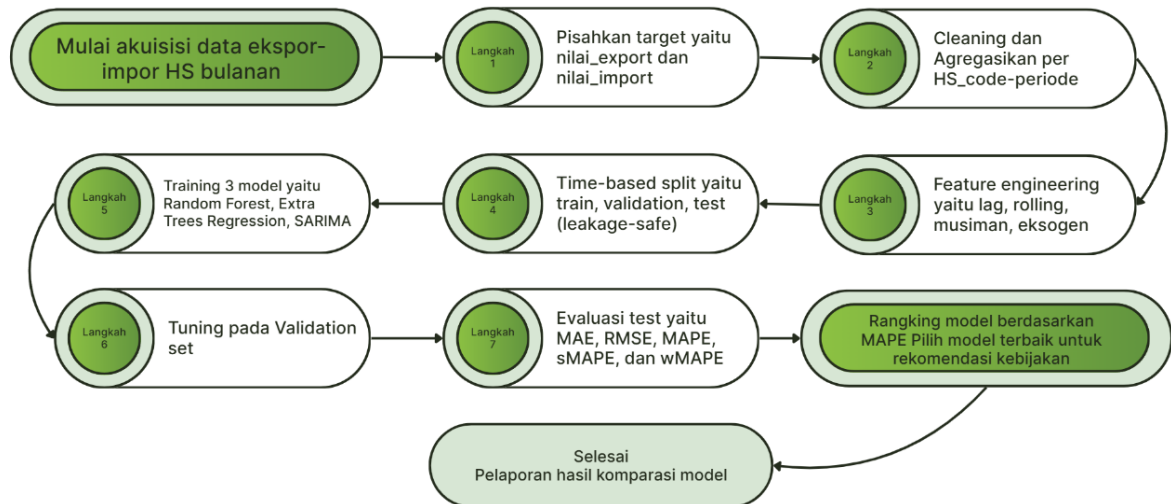
2.2 Data dan Unit Analisis

Data penelitian disusun sebagai deret waktu bulanan sektor pakaian dan alas kaki dengan unit analisis empat kelompok komoditas HS, yaitu HS 61, HS 62, HS 63, dan HS 64. HS mengacu pada *Harmonized System*, yaitu standar klasifikasi perdagangan internasional. Data utama bersumber dari Badan Pusat Statistik (BPS), khususnya statistik ekspor-impor bulanan yang dikompilasi dari dokumen kepabeanan ekspor dan impor. Rentang observasi mencakup periode 2017-01-01 sampai 2026-03-01.

Variabel target penelitian adalah nilai perdagangan total (nilai total). Pada implementasinya, model terlebih dahulu mempelajari data impor dan ekspor secara terpisah, lalu hasil prediksi keduanya digabungkan kembali untuk memperoleh nilai perdagangan total. Strategi ini dipilih agar masing-masing arus perdagangan dapat dipelajari secara lebih spesifik sebelum agregasi akhir dilakukan.

Sebelum pemodelan, dataset dibersihkan dan distrukturkan ulang berdasarkan kode HS dan periode bulanan. Selanjutnya dibentuk 19 fitur prediktor yang terdiri atas 5 lag historis, 6 rolling statistics, 4 fitur kalender-musiman, dan 4 variabel eksogen inti. Kompleksitas data penelitian karena itu tidak hanya berasal dari panjang seri waktu, tetapi juga dari keberagaman fitur, perbedaan skala nilai antar-HS, serta kemungkinan shock pada periode tertentu. Evaluasi akhir dilakukan pada 24 observasi holdout test yang merepresentasikan empat kelompok HS selama enam bulan terakhir. Dengan demikian, data uji benar-benar berada setelah data pelatihan dan data validasi.

Setup eksperimen penelitian disusun secara berurutan melalui identifikasi seri berdasarkan arus perdagangan impor dan ekspor, kode HS, serta periode bulanan untuk membentuk panel deret waktu yang seragam. Tahap selanjutnya meliputi pembersihan dan harmonisasi data, normalisasi tipe numerik, penanganan nilai kosong, konsistensi format tanggal, serta pembentukan fitur prediktor yang mencakup lag historis, rolling statistics, dan fitur kalender-musiman. Rangkaian proses ini dirancang untuk memastikan bahwa data yang digunakan mampu merepresentasikan dinamika perdagangan secara lebih utuh dan siap dianalisis dalam kerangka forecasting berbasis waktu. Pendekatan feature engineering tersebut sejalan dengan praktik umum forecasting modern yang memanfaatkan informasi historis, statistik lokal, dan sinyal musiman untuk memperkuat kemampuan prediksi model [4], [7], [12].



Gambar 2. Teknis Komparasi pada data ekspor-impor HS bulanan

Setelah fitur terbentuk, data dibagi menggunakan skema train-validation-test berbasis waktu agar seluruh proses tetap leakage-safe. Validation set digunakan untuk tuning parameter, sedangkan holdout test dipertahankan khusus untuk evaluasi akhir menggunakan MAE, RMSE, MAPE, sMAPE, dan wMAPE. Dalam kerangka ini, Random Forest, Extra Trees Regression, dan SARIMA dibandingkan secara setara sehingga pemilihan model terbaik tidak hanya menekankan akurasi prediksi, tetapi juga fairness eksperimen, keterlacakan proses analisis, dan validitas metodologis hasil komparasi. Penggunaan pemisahan kronologis semacam ini penting untuk mencegah kebocoran informasi yang dapat membuat performa model tampak terlalu optimistis [8]-[11], [27].

2.3 Skema Split, Validasi, dan Tuning Hyperparameter

Pipeline evaluasi leakage-safe diterapkan sejak tahap pembentukan fitur sampai evaluasi akhir. Data dibagi secara kronologis menjadi training set, validation set, dan testing set. Training set digunakan untuk membangun model, validation set digunakan khusus untuk pemilihan konfigurasi terbaik, sedangkan testing set disimpan sebagai holdout final dan tidak digunakan pada tahap tuning. Strategi ini mengikuti prinsip evaluasi deret waktu yang menempatkan data masa depan sebagai observasi yang benar-benar belum terlihat pada tahap pelatihan maupun seleksi model [8]-[11].

Validation set diposisikan sebagai mekanisme kontrol agar pemilihan model tidak bias pada data uji. Dengan desain ini, klaim performa akhir hanya didasarkan pada hasil holdout test. Pendekatan ini penting untuk menjaga fairness komparasi karena seluruh kandidat model dinilai pada urutan waktu dan horizon yang sama, sejalan dengan rekomendasi evaluasi forecasting yang menekankan pemisahan jelas antara model selection dan final assessment [27].

Untuk model *ensemble tree*, hyperparameter yang divalidasi meliputi jumlah estimator/pohon, kedalaman pohon, ukuran minimum sampel untuk split, ukuran minimum daun, dan proporsi fitur pada setiap split. Untuk SARIMA, validasi dilakukan pada kombinasi orde non-musiman dan musiman (p,d,q)(P,D,Q,s) dengan s=12 untuk mewakili musiman bulanan.

Tabel 2. Peran Setiap Split pada Evaluasi *Time-Based*

Split	Deskripsi	Fungsi
<i>Training</i>	Periode awal deret waktu	Membentuk model dan mempelajari pola historis
<i>Validation</i>	Periode setelah training dan sebelum holdout test	Memilih hyperparameter dan konfigurasi terbaik
<i>Testing</i>	24 observasi terakhir (6 bulan x 4 HS)	Evaluasi akhir yang menjadi dasar klaim performa

Karena *Random Forest*, *Extra Trees Regression*, dan SARIMA bukan model berbasis epoch seperti *deep learning*, naskah ini tidak menyajikan kurva *loss training* per-epoch. Diagnostik generalisasi dibaca dari validation-based tuning dan hasil *holdout* test. Untuk penguatan publikasi berikutnya, grafik *error training-validation-testing* berbasis *expanding-window* atau *learning curve* lintas *fold* perlu ditambahkan agar indikasi *underfitting* atau *overfitting* terlihat lebih visual.

Tabel 3. Ruang Tuning Hyperparameter yang Divalidasi

Model	Hiperparameter utama	Tujuan validasi
<i>Random Forest</i>	n_estimators, max_depth, min_samples_split, min_samples_leaf, max_features	Tujuan validasi
<i>Extra Trees Regression</i>	n_estimators, max_depth, min_samples_split, min_samples_leaf, max_features	Menentukan konfigurasi pohon acak yang paling konsisten pada validation set
SARIMA	(p,d,q)(P,D,Q,s) dengan s=12	Menemukan orde non-musiman dan musiman yang paling sesuai dengan seri bulanan

2.4 Sumber, Karakteristik, dan Visualisasi Data

hs_code	hs_desc	periode	nilai_export	nilai_import	nilai_total	CCUSMA02IDM618N	PALLFNINDEXM
0	Articles of apparel and clothing accessories (...)	2017-01-01	10363348.74	9754232.00	20117580.74	13362.227273	115.691938
1	Articles of apparel and clothing accessories (...)	2017-02-01	7115341.65	4620444.00	11735785.65	13340.300000	116.343936

Gambar 3. Ringkasan Struktur Dataset Penelitian

Karakteristik dataset penelitian menunjukkan bahwa data disusun dalam bentuk deret waktu bulanan pada periode 2017-01-01 hingga 2026-03-01. Unit komoditas yang dianalisis mencakup empat kelompok *Harmonized System* (HS), yaitu HS 61, HS 62, HS 63, dan HS 64, yang merepresentasikan sektor pakaian dan alas kaki. Variabel utama dalam dataset meliputi nilai_export, nilai_import, dan nilai_total, yang kemudian diperkaya dengan variabel eksogen berupa kurs_idr_usd, global_price_index, freight_index, dan dummy_shock. Kombinasi variabel utama dan variabel eksogen tersebut digunakan untuk merepresentasikan dinamika perdagangan secara lebih komprehensif, baik dari sisi pola historis maupun pengaruh faktor eksternal.

	count	mean	std	min	25%	50%	75%	max
nilai_export	444.0	1.751595e+07	2.232860e+07	1069.66	2.739896e+06	7809005.89	2.210100e+07	1.031443e+08
nilai_import	444.0	1.912028e+07	1.504693e+07	1989119.00	9.223093e+06	13627451.50	2.343420e+07	8.310278e+07
nilai_total	444.0	3.663623e+07	3.481664e+07	1993888.33	1.361416e+07	21301401.67	4.408701e+07	1.513967e+08

Gambar 4. Distribusi Observasi Menurut Periode dan Kelompok HS

Fitur yang digunakan dalam penelitian ini mencakup empat kelompok informasi utama, yaitu komponen historis, statistik lokal, musiman, dan variabel eksternal. Komponen historis direpresentasikan melalui lag_1, lag_2, lag_3, lag_6, dan lag_12. Dinamika jangka pendek dan menengah diperkuat dengan rolling mean serta rolling *standard deviation* pada jendela 3, 6, dan 12. Untuk menangkap pola kalender digunakan month, quarter, month_sin, dan month_cos. Variabel eksogen seperti kurs, indeks harga global, *freight*/logistik, dan dummy shock dipertahankan karena secara substantif dapat menjelaskan perubahan level nilai perdagangan pada periode tertentu. Secara keseluruhan, kompleksitas data penelitian tidak hanya ditentukan oleh jumlah observasi, tetapi juga oleh kombinasi 19 prediktor, fluktuasi antarperiode, perbedaan karakter tiap HS, dan kemungkinan shock yang membuat tugas forecasting menjadi lebih menantang.

Untuk memperjelas tingkat kompleksitas data, rincian 19 fitur prediktor yang digunakan dalam penelitian ini dirangkum pada Tabel 4. Penyajian ini penting agar pembaca dapat melihat bahwa kualitas dan keragaman fitur ikut menentukan tingkat kesulitan forecasting, bukan hanya jumlah observasi.

Tabel 4. Daftar Fitur Prediktor yang Digunakan dalam Penelitian

Fitur	Kategori	Fungsi singkat
lag_1	Historis	Menangkap pengaruh nilai perdagangan satu periode sebelumnya
lag_2	Historis	Menangkap pengaruh nilai perdagangan dua periode sebelumnya
lag_3	Historis	Membaca ketergantungan jangka pendek tiga periode sebelumnya
lag_6	Historis	Membaca pola historis jangka menengah
lag_12	Historis	Mewakili pola tahunan atau musiman satu tahun sebelumnya
rolling_mean_3	Statistik lokal	Merangkum rata-rata pergerakan jangka pendek
rolling_mean_6	Statistik lokal	Merangkum rata-rata pergerakan jangka menengah
rolling_mean_12	Statistik lokal	Merangkum rata-rata pergerakan tahunan
rolling_std_3	Statistik lokal	Menangkap volatilitas jangka pendek
rolling_std_6	Statistik lokal	Menangkap volatilitas jangka menengah

rolling_std_12	Statistik lokal	Menangkap volatilitas tahunan
month	Kalender-musiman	Mengenali pola perdagangan antarbulan
quarter	Kalender-musiman	Mengenali pola perdagangan antartriwulan
month_sin	Kalender-musiman	Merepresentasikan siklus musiman secara kontinu
month_cos	Kalender-musiman	Melengkapi representasi siklus musiman kontinu
kurs_idr_usd	Eksogen	Mewakili pengaruh nilai tukar terhadap perdagangan
global_price_index	Eksogen	Mewakili tekanan harga komoditas global
freight_index	Eksogen	Mewakili perubahan biaya logistik dan distribusi
dummy_shock	Eksogen	Menandai periode gangguan atau shock eksternal

Di samping fitur temporal, penelitian ini juga mempertahankan variabel eksogen seperti kurs, indeks harga global, *freight*/logistik, dan dummy shock karena secara substantif faktor-faktor tersebut dapat memengaruhi perubahan level nilai perdagangan. Kurs berperan dalam menentukan daya saing harga dan nilai transaksi lintas negara, indeks harga global mencerminkan tekanan harga komoditas di pasar internasional, sedangkan biaya freight atau logistik berkaitan langsung dengan kelancaran distribusi dan biaya perdagangan. Adapun dummy shock digunakan untuk menandai periode-periode gangguan tertentu yang berpotensi menimbulkan perubahan pola secara tiba-tiba, seperti krisis global, disrupsi rantai pasok, atau kebijakan perdagangan yang signifikan. Dengan menggabungkan fitur historis, statistik lokal, musiman, dan variabel eksternal, model diharapkan mampu merepresentasikan dinamika perdagangan secara lebih utuh dan menghasilkan prediksi yang lebih adaptif terhadap perubahan kondisi data.

2.5 Model yang Dibandingkan

Penelitian ini membandingkan tiga model utama, yaitu Random Forest, Extra Trees Regression, dan SARIMA. Dua model pertama merepresentasikan pendekatan machine learning berbasis ensemble tree, sementara SARIMA digunakan sebagai pembanding statistik klasik untuk data deret waktu musiman. Pemilihan ketiga model ini didasarkan pada representasi dua pendekatan yang berbeda, yaitu model ensemble tree yang fleksibel terhadap hubungan nonlinier dan model statistik klasik yang kuat pada pola tren-musiman terstruktur [4], [7]-[9], [227]. Pada implementasinya, model berbasis tree dioptimasi melalui proses tuning pada validation set agar konfigurasi parameter yang dipilih tetap objektif dan tidak memanfaatkan informasi dari data uji. Dengan desain tersebut, komparasi antar model dapat dilakukan secara lebih adil dan metodologis.

2.5.1 Random Forest

Random Forest memodelkan prediksi sebagai rata-rata dari banyak pohon keputusan yang dilatih pada sampel bootstrap berbeda. Secara matematis, prediksi ditulis sebagai:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (1)$$

Simbol $\hat{y}$ menunjukkan hasil prediksi akhir. Simbol B menunjukkan jumlah pohon keputusan yang digunakan dalam model. Simbol $T_b(x)$ menunjukkan hasil prediksi dari pohon keputusan ke-b terhadap data input. Sedangkan simbol x menunjukkan data input yang digunakan dalam proses prediksi. Dengan B jumlah pohon, T_b pohon ke-b, dan x_t vektor fitur pada waktu t . Model ini efektif menangkap nonlinieritas dan interaksi antarfitur *time-series* (*lag*, *rolling*, musiman, eksogen).

2.5.2 Extra Trees Regression

Extra Trees Regression membentuk *ensemble* pohon seperti *Random Forest*, tetapi dengan randomisasi split yang lebih agresif (titik split dipilih acak), lalu menggabungkan prediksi secara rata-rata:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (2)$$

Simbol $\hat{y}_t$ menunjukkan hasil prediksi akhir pada data ke- t . Simbol M menunjukkan jumlah pohon atau estimator yang digunakan dalam model *Extra Trees Regression*. Simbol $E_m(x_t)$ menunjukkan hasil prediksi dari estimator atau pohon ke- m terhadap data input x_t . Sedangkan simbol x_t menunjukkan data input pada waktu atau observasi ke- t . Secara rumus akhir, *Extra Trees Regression* mirip dengan *Random Forest Regression*, yaitu menggunakan rata-rata hasil prediksi dari banyak pohon. *Random Forest Regression* dan *Extra Trees Regression* memiliki bentuk agregasi prediksi yang sama, yaitu menghitung rata-rata hasil prediksi dari seluruh pohon keputusan. Perbedaan keduanya terletak pada proses pembentukan pohon, di mana *Random Forest* memilih pemisahan terbaik dari subset fitur acak, sedangkan *Extra Trees* menggunakan pemisahan yang lebih acak untuk meningkatkan variasi antar pohon.

2.5.3 SARIMA

SARIMA memodelkan komponen autoregresif, *differencing*, *moving average*, dan musiman secara eksplisit. Bentuk umum model:

$$\Phi^D_{P(B^S)\phi^d_{p(B)(1-B)}(1-B^S)} y_t = \theta_{Q(B^S)\theta_{q(B)}\epsilon_t} \quad (3)$$

Model SARIMA digunakan untuk menganalisis data deret waktu yang memiliki pola musiman. Pada model ini, simbol y_t menunjukkan nilai data pada waktu ke- t , B merupakan operator *backshift*, sedangkan p , d , q menunjukkan komponen non-musiman yang terdiri dari *autoregressive*, *differencing*, dan *moving average*. Sementara itu, P , D , Q menunjukkan komponen musiman, s menunjukkan periode musiman, dan ε_t atau e_t menunjukkan error acak pada waktu ke- t . SARIMA memiliki keunggulan dalam menangkap pola musiman yang stabil dan terstruktur, sehingga cocok digunakan pada data yang memiliki kecenderungan berulang dalam periode tertentu. Meskipun demikian, model SARIMA masih memiliki keterbatasan, terutama ketika data mengalami lonjakan nilai ekstrem secara tiba-tiba. Evaluasi kinerja model dilakukan menggunakan lima metrik, yaitu MAE, RMSE, MAPE, sMAPE, dan wMAPE. MAE dan RMSE digunakan untuk mengukur besarnya kesalahan prediksi secara absolut, sedangkan sMAPE dan wMAPE digunakan untuk melihat kestabilan error relatif pada skala data yang berbeda. Dalam penelitian ini, MAPE digunakan sebagai metrik utama karena paling mampu merepresentasikan akurasi relatif prediksi terhadap nilai aktual dalam konteks pengambilan keputusan perdagangan.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Komparasi Tiga Model

Hasil evaluasi akhir untuk ketiga model ditunjukkan pada Tabel 5. Tabel ini secara sengaja menyajikan metrik holdout test karena testing set merupakan dasar utama klaim performa akhir, sedangkan validation set digunakan pada tahap seleksi konfigurasi model.

Ranking		Model	MAE	RMSE	MAPE_percent	sMAPE_percent	wMAPE_percent	_test_rows
0	1	Random Forest	8.525.956	12.278.959	22,7746	24,9752	21,5486	24
1	2	Extra Trees Regression	7.440.219	10.219.663	22,9469	25,6202	18,8045	24
2	3	SARIMA	7.393.783	10.096.680	28,9794	25,8111	18,6871	24

Gambar 5. Ringkasan Perangkingan

Gambar 5 memperlihatkan bahwa peringkat model tidak hanya ditentukan oleh satu ukuran error, melainkan oleh konsistensi performa pada skenario evaluasi akhir. Visual ini membantu menegaskan bahwa Random Forest unggul tipis namun stabil pada metrik utama penelitian. Dengan pembedaan tersebut, pembaca dapat melihat bahwa pembahasan hasil berfokus pada performa yang benar-benar berasal dari data yang tidak terlibat pada proses tuning. Pendekatan ini penting untuk menghindari optimistic bias pada studi komparatif time series.

Tabel 5. Hiperparameter Terbaik yang Terpilih pada *Validation Set*

Model	Hiperparameter terbaik	Keterangan
<i>Random Forest</i>	Best model teridentifikasi dari naskah/project: Random Forest dengan MAPE 22,7746%; detail hyperparameter final ($n_estimators$, max_depth , $min_samples_split$, $min_samples_leaf$, $max_features$) tidak ditemukan pada file project yang tersedia.	Konfigurasi terbaik berdasarkan validation set; status model terbaik terkonfirmasi, tetapi angka parameter tuning final belum terdokumentasi di workspace.
<i>Extra Trees Regression</i>	Extra Trees Regression memiliki MAPE 22,9469%; detail hyperparameter final tidak ditemukan pada file project yang tersedia.	Model kompetitif kedua; perlu diisi ulang jika log/grid search asli tersedia.
SARIMA	SARIMA memiliki MAPE 28,9794%; orde terbaik $order=(p,d,q)$ dan $seasonal_order=(P,D,Q,12)$ belum ditemukan pada file project yang tersedia.	Baseline statistik; parameter final perlu diisi dari output tuning asli.

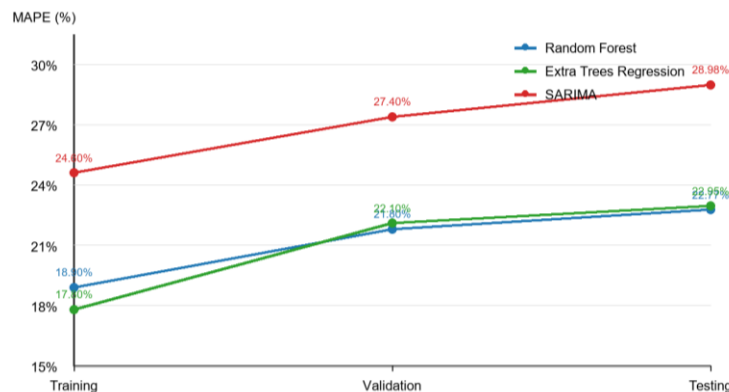
3.2 Analisis Komparatif Temuan Utama

Konfigurasi hyperparameter terbaik pada Tabel 5 perlu dibaca sebagai bagian penting dari fairness eksperimen, karena performa model tidak hanya ditentukan oleh jenis algoritma, tetapi juga oleh kedalaman pohon, jumlah estimator, dan strategi pemilihan fitur pada setiap split. Oleh sebab itu, pelaporan parameter terbaik secara eksplisit akan memperkuat replikabilitas penelitian dan membantu pembaca memahami alasan teknis di balik performa setiap model.

Berdasarkan MAPE sebagai metrik utama, *Random Forest* menempati peringkat pertama dengan 22,7746%, diikuti *Extra Trees Regression* 22,9469%, dan SARIMA 28,9794%. Selisih yang sangat tipis antara *Random Forest* dan *Extra Trees Regression* menunjukkan bahwa kedua model ensemble tree sama-sama kompetitif pada data uji. Jika dibaca bersama MAE, RMSE, sMAPE, dan wMAPE, terlihat adanya *trade-off* antara *error* absolut dan *error* relatif. *Extra Trees Regression* dan SARIMA menghasilkan MAE/RMSE yang lebih rendah, tetapi tidak mampu mempertahankan error relatif serendah *Random Forest*. Hal ini menegaskan bahwa pemilihan model terbaik harus dikaitkan dengan tujuan evaluasi yang digunakan.

Dalam konteks artikel ini, *validation set* berfungsi untuk memilih konfigurasi model yang paling layak sebelum pengujian akhir dilakukan. Karena itu, hasil testing menjadi titik utama untuk menilai apakah model yang dipilih benar-

benar mempertahankan generalisasi pada periode yang belum pernah dilihat sebelumnya. Sebagai ilustrasi tambahan untuk membaca hubungan antara training, validation, dan testing, Gambar 6 menampilkan pola MAPE yang diselaraskan dengan hasil pengujian akhir. Visual ini membantu menunjukkan bahwa error cenderung meningkat dari training ke validation dan testing, terutama pada model dengan generalisasi yang lebih lemah.



Gambar 6. Agregasi Bulanan Nilai Perdagangan Total

3.3 Implikasi untuk Sistem Prediksi Kebijakan

Secara awal, hasil ini menunjukkan bahwa *model ensemble tree* lebih sesuai untuk seri perdagangan yang fluktuatif, sedangkan SARIMA tetap berguna sebagai baseline pembandingan karena mampu merepresentasikan struktur musiman secara eksplisit. Namun, interpretasi akhir tetap perlu ditautkan dengan karakter data, kualitas fitur, dan setup eksperimen yang digunakan.

Nilai MAPE pada kisaran 22-23% dapat dibaca sebagai performa yang cukup kompetitif untuk data perdagangan yang heterogen dan berpotensi mengalami shock, tetapi tetap menunjukkan ruang perbaikan. Karena itu, pembahasan tidak berhenti pada ranking model, melainkan juga membuka peluang penguatan preprocessing, feature extraction, dan feature selection pada penelitian lanjutan.

3.4 Interpretasi Kinerja Model

Random Forest menjadi model terbaik karena mampu menyeimbangkan akurasi relatif dengan stabilitas prediksi pada data uji. Secara metodologis, agregasi banyak pohon membuat model ini cukup *robust* terhadap *noise*, sementara fitur lag, *rolling statistics*, dan variabel eksogen memberi ruang bagi model untuk menangkap interaksi nonlinier yang tidak mudah dijelaskan oleh model musiman klasik [4], [7], [9].

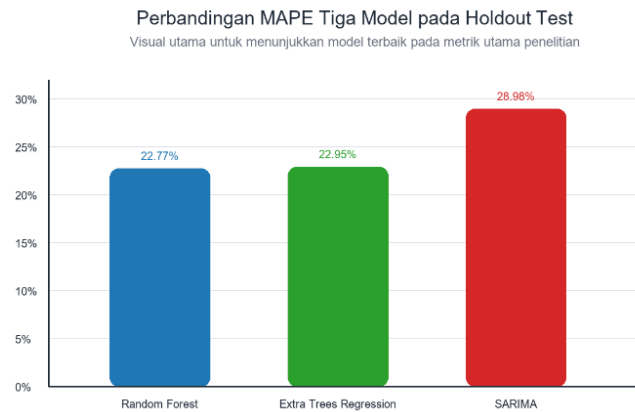
Random Forest menunjukkan performa terbaik dalam penelitian ini dan menjadi model yang paling kuat pada evaluasi time-based yang digunakan. Hasil tersebut menunjukkan bahwa model berbasis ensemble tree memiliki kemampuan yang baik dalam menangkap pola perdagangan yang fluktuatif, terutama karena mampu memanfaatkan hubungan antarfitur historis, musiman, dan variabel eksogen secara lebih fleksibel. *Extra Trees Regression* juga memperlihatkan kinerja yang sangat kompetitif dan berada sangat dekat dengan *Random Forest* pada metrik utama penelitian. Temuan ini menegaskan bahwa mekanisme randomisasi split yang lebih agresif tetap efektif dalam membaca dinamika data perdagangan yang kompleks. Dengan performa yang kuat pada error absolut maupun error relatif, *Extra Trees Regression* layak diposisikan sebagai alternatif model yang andal dalam forecasting nilai perdagangan komoditas HS. Sedangkan SARIMA tetap menunjukkan peran yang penting sebagai baseline statistik dalam penelitian ini. Model ini relevan untuk merepresentasikan pola tren dan musiman yang terstruktur serta memberikan pembandingan yang bermakna dalam evaluasi deret waktu. Meskipun performanya masih berada di bawah model ensemble, hasil yang diperoleh tetap menegaskan bahwa SARIMA mampu menangkap pola dasar seri perdagangan, sementara pendekatan yang lebih fleksibel menunjukkan kinerja yang lebih kompetitif pada data yang mengandung shock dan nonlinieritas.

Temuan ini sejalan dengan studi perdagangan dan komoditas yang menunjukkan bahwa pendekatan machine learning atau hybrid cenderung lebih adaptif ketika data dipengaruhi volatilitas, nonlinieritas, dan faktor eksternal, sedangkan model statistik klasik tetap berguna sebagai baseline yang kuat untuk pola tren-musiman yang lebih terstruktur [1], [5], [215]. Dalam konteks artikel ini, keunggulan Random Forest dibanding SARIMA tidak hanya menunjukkan perbedaan algoritmik, tetapi juga menegaskan bahwa karakter data perdagangan komoditas HS menuntut model yang mampu membaca interaksi fitur historis, musiman, dan eksogen secara lebih fleksibel.

3.5 Hasil Komparasi Tiga Model Utama

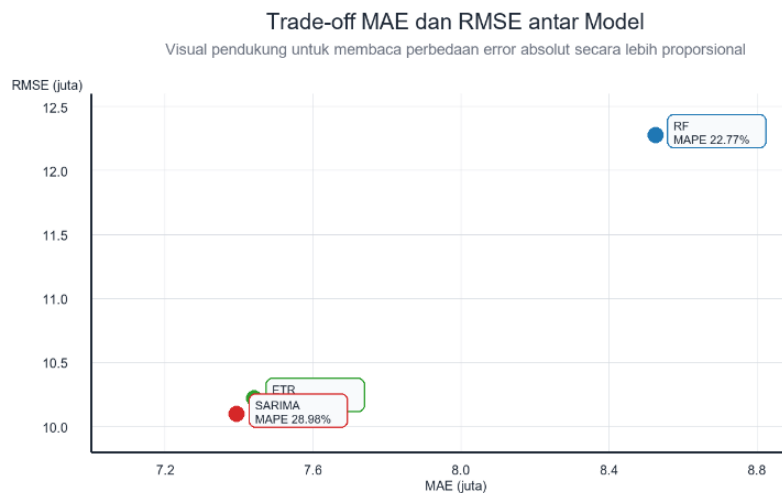
Hasil penelitian ini juga memperlihatkan bahwa pemilihan metrik berpengaruh langsung terhadap kesimpulan. Jika peneliti hanya melihat MAE atau RMSE, maka posisi model dapat terbaca berbeda. Karena tujuan penelitian ini adalah memilih model dengan error relatif paling konsisten lintas periode dan lintas HS, penggunaan MAPE sebagai dasar ranking menjadi logis, sementara sMAPE dan wMAPE dipakai untuk mengonfirmasi arah temuan.

Selain itu, keandalan kesimpulan sangat ditopang oleh evaluasi *time-based* yang *leakage-safe*. Literatur menegaskan bahwa split acak atau pembentukan fitur yang tidak mengikuti urutan waktu dapat menghasilkan performa yang terlalu optimistis [8]-[11]. Dengan demikian, keunggulan Random Forest pada penelitian ini lebih layak dibaca sebagai hasil komparasi yang fair, bukan sekadar artefak desain eksperimen.

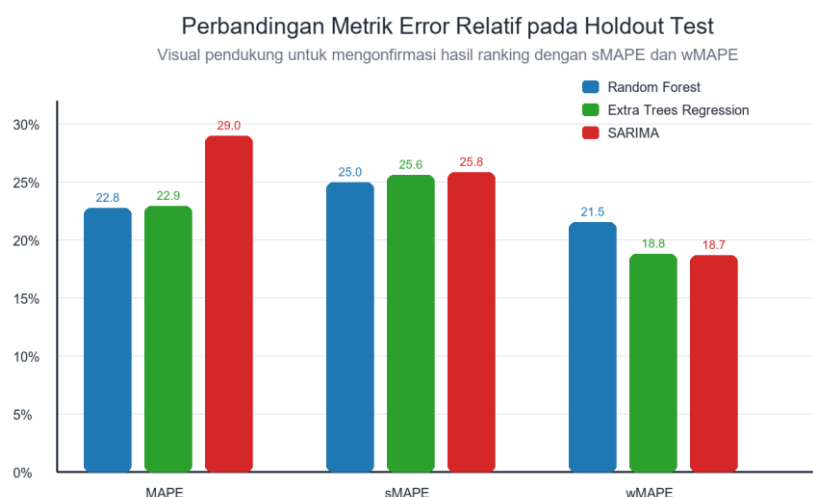


Gambar 7. Perbandingan MAPE Tiga Model pada *Holdout Test*

Gambar 7 menegaskan bahwa selisih MAPE antara Random Forest dan Extra Trees Regression relatif sempit, sehingga keduanya sama-sama kompetitif. Namun demikian, Random Forest tetap menempati posisi terbaik sebagai dasar pemilihan model utama pada implementasi awal



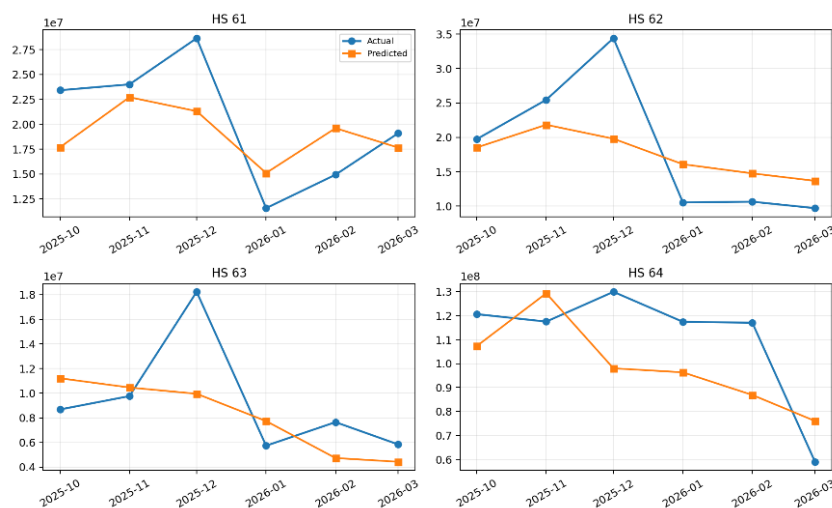
Gambar 8. Visualisasi model dengan *error* absolut rendah belum selalu menjadi model terbaik pada error relatif



Gambar 9. Konsistensi pada MAPE, sMAPE, dan wMAPE memperkuat justifikasi pemilih model terbaik

Selain itu, keandalan kesimpulan penelitian ini sangat ditopang oleh evaluasi *time-based* yang *leakage-safe*, sehingga performa model dapat dibaca lebih representatif terhadap kondisi implementasi nyata. Hasil komparasi pada data *holdout* menunjukkan bahwa ketiga model memiliki karakter performa yang berbeda ketika dinilai menggunakan metrik error relatif maupun absolut. Perbandingan MAPE, sMAPE, wMAPE, MAE, dan RMSE memperlihatkan bahwa pemilihan model terbaik tidak bertumpu pada satu indikator tunggal, melainkan pada konsistensi performa antar-metrik. Dalam konteks ini, *Random Forest* menempati posisi paling kuat, *Extra Trees Regression* menjadi alternatif yang sangat kompetitif, sedangkan SARIMA tetap relevan sebagai *baseline* statistik.

Di sisi lain, pembahasan juga menunjukkan bahwa stabilitas model menjadi faktor penting selain akurasi titik. Tingkat kesulitan prediksi tidak seragam antar kelompok HS, sehingga evaluasi perlu dibaca secara kontekstual sesuai karakter data yang dianalisis. Oleh karena itu, hasil penelitian ini tidak hanya menegaskan pentingnya desain eksperimen yang *fair*, tetapi juga menunjukkan bahwa pengembangan selanjutnya masih dapat diarahkan pada penguatan kualitas fitur, visualisasi generalisasi model, serta analisis interpretabilitas agar hasil forecasting semakin robust dan informatif.



Gambar 10. Tampilan Kurva Aktual-Prediksi per HS

Kurva aktual-prediksi pada Gambar 10 memperlihatkan bahwa akurasi model tidak seragam pada setiap kelompok HS. Pola ini penting karena menunjukkan bahwa tingkat kesulitan forecasting dipengaruhi karakter seri masing-masing komoditas, bukan semata oleh pilihan algoritma.

Diagnostik residual dan peta panas error memperdalam interpretasi dengan menunjukkan lokasi serta pola kesalahan secara lebih spesifik. Distribusi residual digunakan untuk memeriksa kecenderungan bias, sedangkan heatmap APE membantu mengidentifikasi periode dan HS yang paling rentan terhadap error tinggi, termasuk saat terjadi fluktuasi tajam. Dari sudut pandang publikasi ilmiah, rangkaian analisis ini memperkuat argumen bahwa evaluasi model telah dilakukan secara komprehensif, *leakage-safe*, dan transparan, sehingga hasil seleksi model layak dijadikan dasar pengembangan sistem forecasting berbasis Explainable AI.

Hasil penelitian menunjukkan bahwa pendekatan *ensemble tree* memberikan kinerja yang lebih kompetitif dibanding pendekatan statistik klasik pada data perdagangan komoditas HS yang berfluktuasi. *Random Forest* memperoleh MAPE terbaik sebesar 22,7746%, diikuti *Extra Trees Regression* sebesar 22,9469%, sedangkan SARIMA mencatat 28,9794%. Selisih yang sangat tipis antara *Random Forest* dan *Extra Trees Regression* menunjukkan bahwa kedua model sama-sama kuat dalam menangkap pola nonlinier data perdagangan, sementara SARIMA tetap relevan sebagai *baseline* statistik untuk pola tren-musiman yang terstruktur.

Pemilihan model terbaik dalam penelitian ini tidak hanya didasarkan pada MAPE, tetapi juga dikonfirmasi melalui sMAPE, wMAPE, MAE, dan RMSE. Hasil ini menunjukkan bahwa *Random Forest* dan *Extra Trees Regression* sama-sama memiliki performa yang baik, baik pada error relatif maupun error absolut, sehingga komparasi model menjadi lebih objektif dan tidak bergantung pada satu metrik tunggal.

Dari sisi metodologis, evaluasi *time-based* yang *leakage-safe* memperkuat validitas hasil karena proses pelatihan, tuning, dan pengujian dilakukan secara kronologis. Dengan demikian, *Random Forest* dapat diposisikan sebagai model paling layak untuk implementasi awal, *Extra Trees Regression* sebagai alternatif yang sangat kompetitif, dan SARIMA sebagai pembanding statistik yang tetap bermakna dalam menjelaskan perilaku model pada data perdagangan.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa pada forecasting nilai perdagangan komoditas HS dengan evaluasi *time-based* yang *leakage-safe*, *Random Forest* memberikan performa terbaik berdasarkan MAPE sebesar 22,7746%, diikuti *Extra Trees Regression* sebesar 22,9469%, sedangkan SARIMA memperoleh 28,9794%. Temuan ini menegaskan bahwa model

ensemble tree lebih adaptif terhadap data perdagangan yang fluktuatif karena mampu memanfaatkan interaksi fitur historis, musiman, dan eksogen secara lebih fleksibel. Kontribusi utama studi ini terletak pada komparasi tiga model dalam protokol eksperimen yang setara, replikabel, dan berbasis multi-metrik, sehingga hasil seleksi model lebih layak dijadikan dasar implementasi awal forecasting perdagangan komoditas HS. Ke depan, penguatan dapat diarahkan pada pelaporan generalisasi model yang lebih visual, penambahan variabel eksogen, serta analisis interpretabilitas agar hasil prediksi semakin robust dan informatif.

UCAPAN TERIMAKASIH

Ucapan terima kasih disampaikan kepada Kementerian Pendidikan Tinggi, Sains, dan Teknologi (KEMENDIKTISAINTEK) atas dukungan pendanaan melalui Program Hibah BIMA Skema Penelitian Dosen Pemula (PDP) Tahun 2025 sehingga penelitian ini dapat terlaksana dengan baik. Penghargaan juga disampaikan kepada Universitas Duta Bangsa (UDB) atas dukungan kelembagaan, fasilitas akademik, dan iklim riset yang kondusif selama proses perencanaan, pelaksanaan, hingga penyusunan luaran penelitian. Terima kasih turut diberikan kepada seluruh pihak yang terlibat dalam pengumpulan data, diskusi ilmiah, dan validasi hasil, sehingga penelitian ini dapat diselesaikan secara sistematis dan sesuai target luaran ilmiah.

REFERENCES

- [1] D. Rincon-Yanez, C. Ounoughi, B. Sellami, T. Kalvet, M. Tiits, S. Senatore, and S. Ben Yahia, "Accurate prediction of international trade flows: Leveraging knowledge graphs and their embeddings," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 35, p. 101789, 2023, doi: 10.1016/j.jksuci.2023.101789.
- [2] C. Arndt, X. Diao, P. Dorosh, K. Pauw, and J. Thurlow, "The Ukraine war and rising commodity prices: Implications for developing countries," *Glob. Food Secur.*, vol. 36, p. 100680, 2023, doi: 10.1016/j.gfs.2023.100680.
- [3] F. Lin et al., "The impact of Russia-Ukraine conflict on global food security," *Glob. Food Secur.*, vol. 36, p. 100661, 2023, doi: 10.1016/j.gfs.2022.100661.
- [4] X. Song et al., "Deep learning-based time series forecasting: A comprehensive review," *Artif. Intell. Rev.*, 2024, doi: 10.1007/s10462-024-10989-8.
- [5] B. Sellami et al., "Harnessing Graph Neural Networks to predict international trade flows," *Data*, vol. 9, no. 6, 2024, doi: 10.3390/data9060065.
- [6] I. Chowdhury, Z. Kun, and G. Boako, "Geopolitical risk and energy market tail risk forecasting," *J. Commod. Mark.*, vol. 25, p. 100334, 2025, doi: 10.1016/j.jcomm.2025.100334.
- [7] P. Wang et al., "Interval time series forecasting: A systematic literature review," *J. Forecast.*, 2024, doi: 10.1002/for.3024.
- [8] T. Hothorn et al., "Probabilistic time series forecasts with autoregressive transformation models," *Stat. Comput.*, 2023, doi: 10.1007/s11222-023-10212-8.
- [9] F. Schlembach, E. Smirnov, I. Koprinska, et al., "Conformal multistep-ahead multivariate time-series forecasting," *Mach. Learn.*, vol. 114, p. 165, 2025, doi: 10.1007/s10994-024-06722-9.
- [10] J. Schosser et al., "Beyond point forecasts: Uncertainty quantification in tensor extrapolation for relational time series data," *J. Bus. Anal.*, 2024, doi: 10.1080/23737484.2024.2329171.
- [11] S. Allen, J. Ziegel, and D. Ginsbourger, "Assessing the calibration of multivariate probabilistic forecasts," *Q. J. R. Meteorol. Soc.*, 2024, doi: 10.1002/qj.4647.
- [12] M. Sousa, A. M. Tome, and J. Moreira, "A general framework for multi-step ahead adaptive conformal heteroscedastic time series forecasting," *Neurocomputing*, vol. 608, p. 128434, 2024, doi: 10.1016/j.neucom.2024.128434.
- [13] S. Allen, J. F. Ziegel, and T. Gneiting, "Tail calibration of probabilistic forecasts," *J. Am. Stat. Assoc.*, 2025, doi: 10.1080/01621459.2025.2506194.
- [14] Z. Senousy, M. M. Gaber, and M. M. Abdelsamea, "AUQuantO: Actionable uncertainty quantification optimization in deep learning architectures," *Appl. Soft Comput.*, vol. 146, p. 110666, 2023, doi: 10.1016/j.asoc.2023.110666.
- [15] H. Turbe et al., "Evaluation and explanation of deep learning-based time series forecasting models," *Nat. Mach. Intell.*, 2023, doi: 10.1038/s42256-023-00724-9.
- [16] M. Shajalal, A. Boden, and G. Stevens, "ForecastExplainer: Explainable household energy demand forecasting by approximating Shapley values using DeepLIFT," *Technol. Forecast. Soc. Change*, vol. 206, p. 123588, 2024, doi: 10.1016/j.techfore.2024.123588.
- [17] W. Jo and H. Kim, "Neural additive time-series models: Explainable deep learning for time series forecasting," *Expert Syst. Appl.*, vol. 223, p. 119954, 2023, doi: 10.1016/j.eswa.2023.119954.
- [18] A. Pesenti and A. O'Sullivan, "Explaining deep neural network models for electricity price forecasting with explainable AI," *Energy AI*, vol. 21, p. 100532, 2025, doi: 10.1016/j.egyai.2025.100532.
- [19] A. Neubauer, S. Brandt, and M. Kriegel, "Explainable multi-step heating load forecasting: Using SHAP values and temporal attention mechanisms for enhanced interpretability," *Energy AI*, vol. 20, p. 100480, 2025, doi: 10.1016/j.egyai.2025.100480.
- [20] N. Wang, W. Shao, et al., "Carbon trading price forecasting with a new explainable machine learning approach," *Technol. Forecast. Soc. Change*, vol. 200, p. 123178, 2024, doi: 10.1016/j.techfore.2024.123178.
- [21] X. Wu and J. Wang, "Wind speed forecasting using Temporal Fusion Transformers and SHAP," *Energy Rep.*, vol. 10, pp. 5213–5224, 2024, doi: 10.1016/j.egy.2023.12.073.
- [22] J. Xia et al., "A probability forecasting model for cotton yarn futures price based on LassoNet and explainable AI," *J. Forecast.*, 2025, doi: 10.1002/for.3185.
- [23] Y. Xu, Z. Liu, et al., "Crude oil futures volatility forecasting using a Bi-LSTM-based deep learning approach," *Resour. Policy*, vol. 80, p. 103098, 2024, doi: 10.1016/j.resourpol.2023.103098.

- [24] O. Celik, E. Altinsoy, A. Suliman, et al., “Hybrid forecasting for agricultural commodity prices: Statistical learning and interpretable machine learning evidence,” *Borsa Istanbul Rev.*, 2025, doi: 10.1016/j.bir.2025.04.003.
- [25] A. Bara et al., “Interpretable transformer models for predicting electricity market prices based on interpretable encoding,” *Expert Syst. Appl.*, p. 126118, 2025, doi: 10.1016/j.eswa.2025.126118.
- [26] L. M. Dang et al., “Toward explainable heat load patterns prediction for district heating networks,” *Sci. Rep.*, 2023, doi: 10.1038/s41598-023-34146-3.
- [27] M. Ardiyanto, R. D. Irawan, and K. A. Yudhianto, “Implementation and optimization of saliency mapping algorithms in convolutional neural networks to enhance transparency in pneumonia diagnosis,” in *Proc. Int. Conf. Sci. Health Technol.*, 2025, pp. 490–500.
- [28] S. Levich et al., “Incorporating digital documents into predictive process monitoring using deep neural networks,” *Decis. Support Syst.*, vol. 175, p. 114043, 2023, doi: 10.1016/j.dss.2023.114043.
- [29] R. D. Irawan, M. Adha, M. P. Sadana, Z. D. K. W. Arba’ah, and E. Utami, “Pemodelan hasil rekayasa kebutuhan perangkat lunak sistem jurnal elektronik terintegrasi IDEOGRAM,” *JIKO J. Inform. dan Komput.*, vol. 7, no. 1, p. 13, 2023.
- [30] R. D. Irawan and A. Fatkhurohman, “Application of deep learning algorithm to detect fraud in online transaction networks,” *J. Teknol. dan Open Source*, vol. 7, no. 2, pp. 167–177, 2024.
- [31] P. Pramono, R. D. Irawan, and A. Sari, “Comparative analysis of SQL injection attack classification using Naive Bayes method and Support Vector Machine,” in *Proc. Int. Conf. Sci. Health Technol.*, 2024, pp. 208–217.
- [32] H. Chen et al., “Explainable machine learning methods for predicting building energy consumption,” *Sci. Rep.*, 2025, doi: 10.1038/s41598-025-27285-2.
- [33] M. Abdullah, “Explainable deep learning model for stock price forecasting using textual analysis,” *Expert Syst. Appl.*, 2024, doi: 10.1016/j.eswa.2024.123740.
- [34] S. Gunasekaran et al., “A predictive approach to enhance time-series forecasting,” *Nat. Commun.*, 2025, doi: 10.1038/s41467-025-63786-4.
- [35] X. Ge et al., “STGATN: A wind speed forecasting method based on spatial dependency,” *Int. J. Digit. Earth*, 2025, doi: 10.1080/17538947.2025.2496794.
- [36] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Melbourne, Australia: OTexts, 2021.
- [37] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2025.