Sistemasi: Jurnal Sistem Informasi
Volume 15, Nomor 11, 2026: 4878-4892
ISSN:2302-8149
e-ISSN:2540-9719

Evaluasi Model *Machine Learning* dan *Ensemble* untuk Prediksi Permintaan Multi-SKU dengan Validasi Deret Waktu

Evaluation of Machine Learning and Ensemble Models for Multi-SKU Demand Forecasting with Time-Series Validation

[1]Sri Devi*, [2]Ida Nurhaida
[1,2]Program Studi Informatika, Fakultas Teknologi dan Desain, Universitas Pembangunan Jaya
[1,2]Jl. Boulevard UPJ, Bintaro Jaya, Kec. Ciputat, Kota Tangerang Selatan, Banten 15413, Indonesia
*e-mail: [1]sri.devi@student.upj.ac.id, [2]ida.nurhaida@upj.ac.id

(***received:*** 14 November 2026, ***revised:*** 20 November 2026, ***accepted:*** 21 November 2026)

Abstrak

Perencanaan persediaan pada gudang distribusi sarana pertanian kerap terkendala fluktuasi pesanan musiman dan tingginya proporsi minggu tanpa transaksi. Penelitian ini mengevaluasi sepuluh metode peramalan—baseline statistik (Naïve, MA-4, ETS, Croston-SBA), Machine Learning (RF, SVR), Deep Learning (LSTM), Hybrid (LSTM+RF), serta Ensemble (Voting, Weighted Average)—berdasarkan pola permintaan multi-SKU dengan validasi deret waktu bebas kebocoran data. Audit pembersihan terhadap 20.051 baris transaksi (Januari 2024–Agustus 2026) memisahkan 195 retur negatif dan 204 duplikasi menjadi 19.652 transaksi bersih mencakup 145 SKU (137 minggu efektif: 109 latih, 28 uji). Klasifikasi Syntetos-Boylan pada data latih mengidentifikasi 7 SKU Smooth (4,83%), 26 Erratic (17,93%), 18 Intermittent (12,41%), 85 Lumpy (58,62%), dan 9 Inactive (6,21%). Pada 5-fold rolling-origin cross-validation (median MASE 0,779–1,242), Weighted Ensemble terpilih untuk Smooth (0,799) dan Lumpy (0,878), sedangkan LSTM terpilih untuk Erratic (0,779) dan Intermittent (0,993). Pada periode uji 136 SKU aktif, model terpilih validasi mencatat median MASE 0,901 (Smooth), 0,797 (Erratic), 0,633 (Intermittent), dan 0,825 (Lumpy), dengan kinerja gabungan 0,804 (Mean 0,992) yang mengungguli 7 dari 10 model tunggal meski sedikit di bawah MA-4 (0,778), ETS (0,780), dan LSTM (0,793). Secara teoretis, temuan ini menegaskan adanya pergeseran performa validasi-ke-uji sehingga model kompleks tidak selalu mengungguli metode statistik parsimonius pada deret waktu sporadis pendek. Penelitian ini berkontribusi menyajikan kerangka evaluasi multi-SKU bebas data leakage yang terintegrasi dengan simulasi Cycle Service Level (CSL), Safety Stock, dan Reorder Point pada dasbor Streamlit.

Kata kunci: *demand forecasting, ensemble learning,* klasifikasi pola permintaan, *random forest,* validasi deret waktu.

Abstract

Inventory planning in agricultural supply distribution warehouses is often constrained by seasonal fluctuations in orders and a high proportion of weeks with zero transactions. This study evaluates ten forecasting methods—statistical baselines (Naïve, MA-4, ETS, and Croston-SBA), machine learning models (RF and SVR), deep learning (LSTM), a hybrid model (LSTM+RF), and ensemble methods (Voting and Weighted Average)—for multi-SKU demand patterns using time series validation designed to prevent data leakage. A data-cleaning audit of 20,051 transaction records from January 2024 to August 2026 identified and separated 195 negative-return records and 204 duplicate records, resulting in 19,652 cleaned transactions covering 145 SKUs and 137 effective weeks (109 training weeks and 28 testing weeks). Based on the Syntetos–Boylan classification applied to the training data, the SKUs were categorized as 7 Smooth (4.83%), 26 Erratic (17.93%), 18 Intermittent (12.41%), 85 Lumpy (58.62%), and 9 Inactive (6.21%). In five-fold rolling-origin cross-validation, with median MASE values ranging from 0.779 to 1.242, the Weighted Ensemble was selected for the Smooth (0.799) and Lumpy (0.878) demand categories, whereas LSTM was selected for the Erratic (0.779) and Intermittent (0.993) categories. During the testing period, among 136 active SKUs, the models selected through validation achieved median MASE values of 0.901 (Smooth), 0.797 (Erratic), 0.633 (Intermittent), and 0.825 (Lumpy). Overall, the selected models achieved a median MASE of 0.804

 ISSN:2302-8149
e-ISSN:2540-9719

(mean: 0.992), outperforming seven of the ten individual models, although their performance was slightly lower than that of MA-4 (0.778), ETS (0.780), and LSTM (0.793). Theoretically, these findings highlight a shift in model performance between validation and testing, indicating that complex models do not necessarily outperform parsimonious statistical methods when applied to short and intermittent time series. This study contributes an integrated, leakage-free multi-SKU evaluation framework that incorporates Cycle Service Level (CSL), Safety Stock, and Reorder Point simulations into a Streamlit dashboard.
Keywords: *demand forecasting, demand pattern classification, ensemble learning, random forest, time-series validation.*

1 Pendahuluan

Pengelolaan persediaan pada gudang distribusi sarana pertanian dan pestisida memegang peranan penting dalam menjaga ketersediaan pasokan lintas wilayah. Pada operasional gudang mitra perusahaan (nama perusahaan disamarkan menjadi PT Agro Sarana Utama sesuai dengan kesepakatan perlindungan kerahasiaan data komersial), pengelolaan 145 jenis produk (Stock Keeping Unit / SKU) dihadapkan pada variasi pola pesanan distributor yang ekstrem antara periode puncak dan jeda musiman. Sebagian produk utama memiliki transaksi pengeluaran rutin setiap minggu, namun mayoritas item pestisida khusus mengalami proporsi minggu tanpa transaksi (zero-demand weeks) di atas 65%. Ketidakpastian estimasi kebutuhan barang keluar ini meningkatkan risiko penumpukan stok lambat gerak (overstock) yang membebani ruang simpan, maupun kekosongan barang (stockout) ketika terjadi lonjakan pesanan serentak [1], [2]. Perencanaan pengadaan yang selama ini mengandalkan rekapitulasi lembar kerja manual dan rata-rata bergerak sederhana (Moving Average) belum mampu beradaptasi terhadap perbedaan perilaku deret waktu antar-SKU [3], [22].

Sejumlah penelitian terdahulu telah mengaplikasikan algoritma Machine Learning seperti Random Forest (RF) dan Support Vector Regression (SVR) maupun jaringan saraf berulang Long Short-Term Memory (LSTM) dalam peramalan rantai pasok [5], [6], [7], [9], [10]. Walaupun demikian, kajian literatur menunjukkan tiga celah metodologis yang jarang ditangani secara simultan. Pertama, evaluasi model umumnya hanya diuji pada satu produk bervolume besar [11] tanpa memetakan matriks karakteristik permintaan multi-SKU (smooth, erratic, intermittent, dan lumpy) [4], [8], [17]. Kedua, penggunaan validasi silang acak (standard k-fold cross-validation) serta penentuan bobot gabungan (ensemble) dari kesalahan data latih berisiko menimbulkan kebocoran informasi masa depan (temporal data leakage) [3], [12], [26]. Ketiga, pelaporan akurasi yang hanya bergantung pada Mean Absolute Percentage Error (MAPE) menghasilkan distorsi persentase ekstrem ketika permintaan aktual mingguan mendekati nol, sehingga diperlukan metrik bebas skala seperti Mean Absolute Scaled Error (MASE) dan Weighted Absolute Percentage Error (WAPE) [3], [15], [26].

Penelitian ini berkontribusi pada evaluasi metode peramalan permintaan berbasis data transaksi riil gudang melalui analisis karakteristik permintaan multi-SKU dan perbandingan model machine learning, deep learning, serta ensemble learning. Evaluasi dilakukan dengan pendekatan validasi deret waktu untuk menjaga independensi temporal antara data pelatihan dan pengujian. Selain membandingkan kesalahan prediksi, penelitian ini menganalisis kesesuaian model terhadap pola permintaan produk yang berbeda dan mengevaluasi pemanfaatan hasil peramalan sebagai masukan dalam perencanaan persediaan. Dengan demikian, kontribusi penelitian diarahkan pada identifikasi metode peramalan yang sesuai dengan karakteristik permintaan dan kebutuhan operasional gudang, bukan semata-mata pemilihan algoritma dengan nilai kesalahan terendah.

2 Tinjauan Literatur

Karakteristik permintaan produk di gudang distribusi tidak selalu mengikuti distribusi normal yang stasioner. Syntetos dan Boylan [4], [8], [17] mengembangkan skema klasifikasi deret waktu berdasarkan parameter Average Demand Interval (ADI) dengan nilai ambang 1,32 dan Squared Coefficient of Variation (CV^2) dengan nilai ambang 0,49 untuk membagi inventori ke dalam empat kuadran: smooth, erratic, intermittent, dan lumpy. Untuk produk berpola intermittent dan lumpy yang didominasi oleh periode permintaan nol, metode statistik khusus seperti pendekatan Croston [16]

 ISSN:2302-8149
e-ISSN:2540-9719

beserta koreksi bias Syntetos-Boylan Approximation (SBA) [4] menjadi pembanding standar yang wajib disertakan di samping metode Naïve dan Exponential Smoothing (ETS) [3], [22].

Dalam ranah pembelajaran mesin, Random Forest bekerja melalui agregasi bootstrap pohon keputusan regresi yang tahan terhadap variansi tinggi pada sampel terbatas [7], [24], sedangkan SVR memetakan hubungan non-linear menggunakan fungsi kernel dengan penalti regularisasi L2 [10], [13], [14]. Pada arsitektur Deep Learning, jaringan LSTM menggunakan mekanisme gerbang memori untuk menangkap ketergantungan sekuensial [9], namun kinerjanya pada deret waktu pendek sangat sensitif terhadap kapasitas parameter dan panjang observasi pelatihan [5], [6], [21]. Bergmeir dan Benítez [12] serta Hewamalage dkk. [26] menegaskan bahwa evaluasi model deret waktu harus menggunakan skema rolling-origin (expanding window) agar pemilihan fitur maupun pembentukan bobot ensemble tidak terkontaminasi oleh data masa depan.

Di samping perbandingan model tunggal, kajian empiris mutakhir menunjukkan bahwa kompleksitas algoritma pembelajaran mesin maupun jaringan saraf tidak selalu menghasilkan akurasi lebih baik daripada metode statistik parsimonius pada deret waktu pendek [19], [22]. Untuk mengatasi ketidakstabilan performa model tunggal pada katalog multi-SKU yang heterogen, literatur terkini merekomendasikan pendekatan kombinasi peramalan (ensemble learning) berbasis rata-rata maupun pembobotan error [5], [25], serta arsitektur hibrida yang memadukan model sekuensial dengan pemodelan residual atau statistik [20]. Pada operasional distribusi dan ritel multi-produk, tantangan utama tidak hanya terletak pada akurasi peramalan permintaan sporadis, melainkan juga pada kepatuhan terhadap protokol evaluasi bebas kebocoran data menggunakan metrik bebas skala [26] serta penyelarasan hasil peramalan dengan parameter pengendalian persediaan seperti Safety Stock, Reorder Point, dan Cycle Service Level (CSL) [18], [23]. Ringkasan posisi penelitian ini dibandingkan studi terdahulu disajikan pada Tabel 1.

Tabel 1 Perbandingan posisi penelitian terhadap studi peramalan permintaan terdahulu

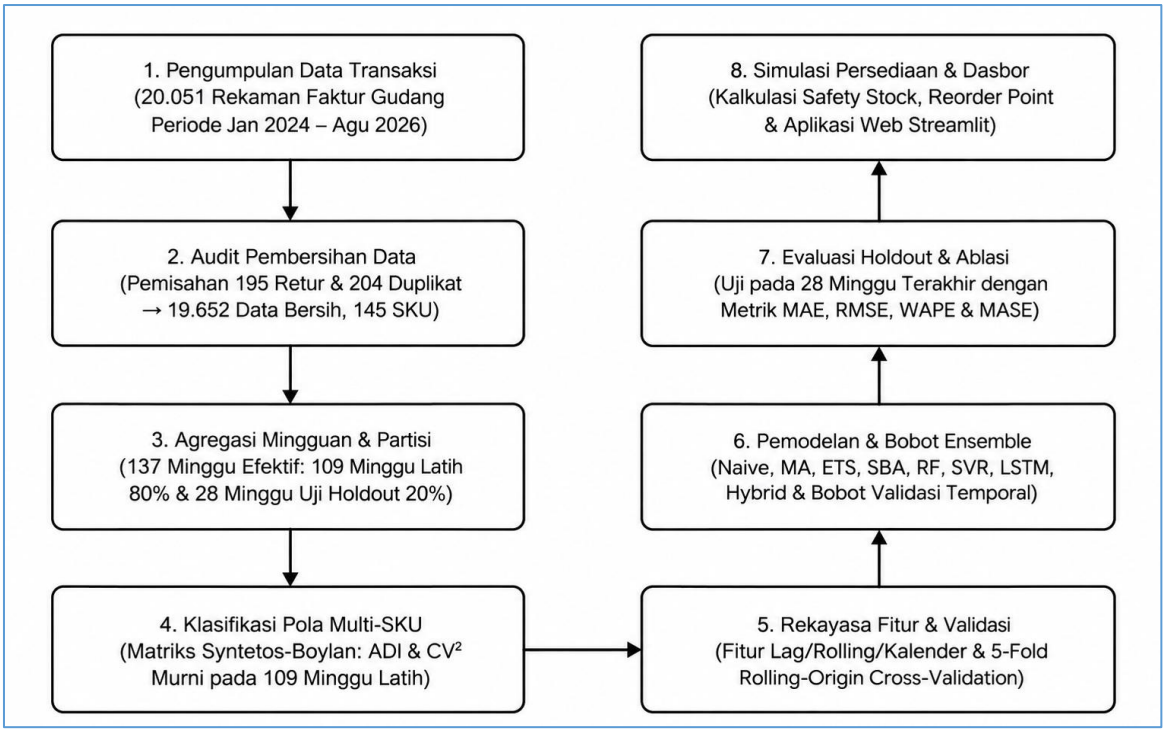
Peneliti & Tahun	Objek & Cakupan SKU	Metode / Model yang Diuji	Skema Validasi Temporal	Metrik Evaluasi	Integrasi Keputusan Persediaan
Carbonneau dkk. [6] (2008)	Rantai pasok manufaktur (Multi-item)	MA, Trend, RNN, SVM, OLS	Train-test split kronologis	MAE, MSE	Tidak mencakup simulasi SS/ROP
Makridakis dkk. [5] (2022)	Retail M5 Competition (Multi-SKU)	Naïve, ETS, ARIMA, LightGBM, Ensemble	Rolling-origin evaluation	WRMSSE, MASE	Fokus pada akurasi peramalan
Alam Khan & Erdoğdu [2] (2025)	Industri pangan (Kajian agregat)	Model berbasis AI & regresi	Pembagian statis	MAPE, RMSE	Konseptual manajemen stok
Palgunadi dkk. [11] (2025)	Perencanaan stok gudang (Single-SKU)	Support Vector Regression (SVR)	Holdout split	MAPE, MSE	Estimasi kebutuhan pengadaan
Penelitian Ini (2026)	Gudang agroindustri (145 SKU; 4 Kuadran ADI-CV²)	Naïve, MA, ETS, Croston-SBA, RF, SVR, LSTM, Hybrid, Ensemble	5-Fold Rolling-Origin CV + 28 Minggu Holdout	MAE, RMSE, WAPE, MASE, MAPE	Simulasi CSL, SS & ROP

3 Metode Penelitian

3.1 Tahapan Penelitian dan Protokol Eksperimen

Tahapan penelitian ini dirancang mengikuti alur sekuensial delapan langkah untuk menjamin reproduktifitas eksperimen dan mencegah kebocoran data temporal (data leakage), sebagaimana ditunjukkan pada Gambar 1. Proses dimulai dari ekstraksi log faktur gudang, audit pembersihan data, agregasi mingguan beserta partisi kronologis, klasifikasi pola permintaan 145 SKU murni pada data pelatihan, rekayasa fitur dan validasi rolling-origin, pelatihan model dasar beserta kalibrasi bobot ensemble dari error validasi, evaluasi akhir pada periode holdout test (28 minggu), hingga simulasi ketersediaan dan rekomendasi Safety Stock (SS).

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 1 Diagram alir tahapan penelitian

3.2 Dataset, Batasan Observasi, dan Audit Pembersihan Data

Data bersumber dari rekaman faktur pengeluaran barang (outbound sales invoice) gudang distribusi PT Agro Sarana Utama periode 3 Januari 2024 hingga 28 Agustus 2026 (rentang 32 bulan atau 139 minggu kalender). Mengingat perusahaan belum mencatat log antrean pesanan yang tertunda atau batal akibat stok kosong (unfulfilled backorders / lost sales), variabel target dalam penelitian ini didefinisikan secara ketat sebagai permintaan pengeluaran barang bruto yang terealisasi (realized gross outbound demand), bukan permintaan pasar tak terbatas (uncensored true demand) maupun arus persediaan bersih (net inventory flow).

Berdasarkan definisi target tersebut, dilakukan audit pembersihan data (data-cleaning audit) terhadap 20.051 baris transaksi awal. Pertama, sebanyak 195 baris (0,97%) yang bernilai negatif atau nol—yang merupakan pencatatan nota retur barang dan koreksi faktur dari transaksi periode sebelumnya—dipisahkan dari dataset sehingga menyisakan 19.856 baris transaksi positif (mencakup 145 SKU aktif setelah 3 kode item koreksi tereliminasi). Apabila retur negatif langsung dikurangkan pada minggu terjadinya retur, volume permintaan aktual pada minggu berjalan akan terdistorsi rendah dan memicu bias under forecasting. Kedua, dilakukan pemeriksaan duplikasi penuh pada keempat atribut (Invoice No., Invoice Date, Item Description, dan Quantity) yang mengidentifikasi 204 baris (1,02%) sebagai duplikasi input sistem pada nomor faktur yang sama persis. Baris transaksi pada faktur yang sama namun memiliki kuantitas berbeda (pengiriman parsial / split batch sebanyak 1.081 baris) tetap dipertahankan karena merupakan pengeluaran fisik barang yang sah. Hasil audit menyisakan 19.652 baris transaksi bersih (2.738 faktur unik) sebagaimana dirangkum pada Tabel 2.

Tabel 2 Hasil audit tahapan pembersihan data (data-cleaning audit)

Tahapan Audit & Penyaringan Data	Jumlah Baris	Proporsi (%)	Jumlah SKU	Justifikasi Metodologis & Dampak pada Target Forecasting
Populasi Baris Transaksi Awal (Raw Data)	20.051	100,00%	148	Seluruh log transaksi gudang periode 3 Jan 2024 – 28 Agu 2026
Pemisahan Transaksi Retur/Koreksi (Quantity <= 0)	195	0,97%	-3	Target adalah gross outbound demand; mencegah distorsi pengurangan permintaan pada minggu retur

 ISSN:2302-8149
e-ISSN:2540-9719

Tahapan Audit & Penyaringan Data	Jumlah Baris	Proporsi (%)	Jumlah SKU	Justifikasi Metodologis & Dampak pada Target Forecasting
Populasi Transaksi Positif (Quantity > 0)	19.856	99,03%	145	Seluruh pengeluaran unit kemasan positif (Min = 1, Max = 32.000)
Eliminasi Duplikasi Sistem (Exact Duplicates)	204	1,02%	0	Menghapus double-entry identik pada faktur yang sama; mempertahankan pengiriman parsial sah
Total Baris Transaksi Bersih (Clean Final)	19.652	98,01%	145	Dataset final valid (2.738 faktur unik) untuk agregasi mingguan

3.3 Agregasi Mingguan, Klasifikasi Multi-SKU (ADI-CV²), dan Rekayasa Fitur

Transaksi harian untuk masing-masing 145 SKU diagregasikan ke dalam kalender mingguan penuh (weekly resampling dengan pengisian nilai nol pada minggu tanpa transaksi), menghasilkan 139 minggu kalender. Hal ini mencakup tiga periode minggu yang bernilai nol (zero-demand weeks) akibat libur operasional pada Hari Raya Idul Fitri, namun berdasarkan analisis data transaksi, lonjakan pemesanan juga teridentifikasi sebagai aktivitas persiapan distributor yang terjadi pada bulan-bulan puncak menjelang estimasi waktu persiapan distribusi pasar, bukan sekadar respons langsung terhadap kalender libur. Setelah pemangkasan dua minggu awal untuk pembentukan fitur waktu mundur, diperoleh 137 minggu observasi efektif yang dibagi secara kronologis menjadi 109 minggu data pelatihan (80%: Januari 2024 – pertengahan Februari 2026) dan 28 minggu data pengujian holdout (20%: akhir Februari 2026 – Agustus 2026).

Penilaian model dilakukan terhadap deret waktu SKU yang memenuhi kriteria berdasarkan informasi pada periode pelatihan. Kriteria kelayakan ditetapkan murni sebelum evaluasi pada data holdout untuk menghindari bias seleksi (post-hoc selection bias) berdasarkan realisasi permintaan masa depan. Produk dengan permintaan nol sepanjang periode pengujian tetap dipertahankan dalam analisis selama denominator metrik (MASE) yang diekstraksi dari data pelatihan bernilai lebih besar dari nol. Klasifikasi pola permintaan dihitung menggunakan parameter Average Demand Interval (ADI) dan Squared Coefficient of Variation (CV^2) menurut Persamaan (1) [4], [8], [17]:

$$ADI = N_train / N_non\text{-}zero\ , \qquad CV^2 = (\sigma_non\text{-}zero / \mu_non\text{-}zero)^2 \tag{1}$$

dengan N_train = 109 minggu, N_non-zero adalah jumlah minggu dengan permintaan lebih besar dari nol pada periode pelatihan, serta σ_non-zero dan μ_non-zero masing-masing menyatakan simpangan baku sampel dan rata-rata unit kemasan pada minggu permintaan non-nol. Mengacu pada ambang batas Syntetos-Boylan (ADI = 1,32 dan CV^2 = 0,49), setiap SKU diklasifikasikan ke dalam kategori Smooth (ADI ≤ 1,32; CV^2 ≤ 0,49), Erratic (ADI ≤ 1,32; CV^2 > 0,49), Intermittent (ADI > 1,32; CV^2 ≤ 0,49), dan Lumpy (ADI > 1,32; CV^2 > 0,49). Sembilan SKU tanpa transaksi pada periode pelatihan (N_non-zero = 0) dipisahkan sebagai kelompok cold-start / no-history (Inactive).

Guna menguji kontribusi variabel input (feature ablation) pada model berbasis pohon dan regresi [24], dirancang tiga konfigurasi fitur waktu mundur yang dihitung dari histori sebelum titik prediksi t: (1) Konfigurasi F1 (Lag-Only) mencakup Lag_1 (y_{t-1}) dan Lag_2 (y_{t-2}); (2) Konfigurasi F2 (Lag + Rolling Statistics) menambahkan Lag_4 (y_{t-4}), rata-rata bergerak 4 minggu (Roll_Mean_4), dan simpangan baku bergerak 4 minggu (Roll_Std_4); serta (3) Konfigurasi F3 (Lag + Rolling + Kalender) menambahkan atribut indeks bulan (1–12) dan kuartal tahunan (1–4). Untuk memastikan perbandingan fitur berjalan adil, seluruh minggu awal yang belum memiliki nilai Lag_4 (t < 5) diisi dengan nilai observasi pertama atau nol (backfill), sehingga ketiga konfigurasi menggunakan jumlah observasi pelatihan dan pengujian yang sama persis tanpa mengubah ukuran dimensi waktu.

3.4 Validasi Rolling-Origin, Horizon Prediksi, dan Formulasi Model

Penalaan hiperparameter dan pemilihan konfigurasi model dilakukan menggunakan validasi deret waktu pada periode pelatihan menggunakan skema 5-fold rolling-origin cross-validation (TimeSeriesSplit) [12], [26]. Setiap fold mempertahankan urutan kronologis, dengan data validasi (18 minggu) selalu berada setelah periode pelatihan. Secara spesifik, Fold 1 melatih indeks observasi [0–18] untuk memvalidasi [19–36], Fold 2 melatih [0–36] memvalidasi [37–54], hingga Fold 5 melatih

 ISSN:2302-8149
e-ISSN:2540-9719

[0–90] memvalidasi [91–108]. Rekayasa fitur (seperti memastikan rolling statistics bergeser menggunakan fungsi shift(1)), transformasi skala dengan MinMaxScaler (yang hanya di-fit pada observasi latih di setiap fold), serta penentuan bobot ensemble dilakukan tanpa memanfaatkan informasi dari periode holdout. Evaluasi akhir menggunakan periode pengujian yang dikunci (indeks 109–136), sehingga hasilnya benar-benar mencerminkan kemampuan model pada data masa depan.

Protokol evaluasi utama menggunakan skema one-step-ahead rolling forecasting (horizon h = 1 minggu) di mana histori di-update menggunakan observasi aktual yang telah terjadi hingga minggu t-1. Pengujian tambahan pada horizon h = 2 dan h = 4 minggu dilakukan murni menggunakan informasi y yang diproyeksikan dari titik t-h, tanpa membocorkan nilai aktual masa depan yang belum tersedia saat prediksi dibuat. Untuk mengatasi kelemahan penentuan bobot berdasarkan training error yang berisiko memberikan bobot berlebih pada model yang mengalami overfitting, bobot model ke-i pada pendekatan Weighted Average Ensemble (w_i) dihitung berdasarkan kebalikan dari Mean Absolute Error pada himpunan validasi temporal lintas lima fold (MAE_{val,i}) sesuai Persamaan (2) [25]:

$$w_i = (1 / MAE_\{val,i\}) / \Sigma_\{j=1\}^\{M\} (1 / MAE_\{val,j\}) \quad (2)$$

Demikian pula pada arsitektur Hybrid (LSTM + RF), model Random Forest kedua dilatih khusus untuk memprediksi komponen residual validasi temporal dari LSTM ($e_\{val,t\} = y_\{val,t\} - \hat{y}_\{LSTM,val,t\}$), bukan residual pada data pelatihan LSTM itu sendiri [20]. Pengujian melibatkan empat metode baseline statistik dan enam skema pembelajaran mesin dan ensemble sebagaimana dirinci pada Tabel 3.

Tabel 3 Spesifikasi metode baseline, model pembelajaran mesin, dan skema ensemble

Kelompok & Nama Model	Konfigurasi Parameter & Arsitektur	Protokol Kalibrasi / Pelatihan	Peran dalam Eksperimen
Baseline Statistik (Naïve & MA-4)	Naïve: y_{t-1}; MA: rata-rata 4 minggu terakhir	Perhitungan langsung tanpa parameter bebas	Batas bawah akurasi deret waktu pendek
Baseline Deret Waktu (ETS & Croston-SBA)	ETS (α=0,35); Croston-SBA: pembaruan terpisah ukuran & interval non-nol (α=0,25)	Kalibrasi penghalusan pada histori pelatihan	Baseline standar khusus pola smooth dan intermittent/lumpy [4], [16]
Random Forest (RF)	n_estimators=100, max_depth=4, min_samples_leaf=4, bobot eksponensial	Penalaan pada 5-fold rolling-origin CV (bobot resensi waktu)	Menangkap hubungan non-linear dengan adaptasi level shift
Support Vector Regression (SVR)	Kernel RBF, C=1,5, epsilon=0,05, gamma='scale', MinMaxScaler	Fitting scaler murni di dalam fold pelatihan	Regresi margin terregularisasi L2 pada sampel terbatas
Long Short-Term Memory (LSTM)	32 unit sekuensial (ReLU), Dropout(0,2), Dense(1), Adam, 45 epoch	Pelatihan berurutan tanpa pengacakan (shuffle=False)	Evaluasi kapasitas jaringan sekuensial pada N=109 minggu
Hybrid (LSTM + Residual RF)	Prediksi utama LSTM + RF (max_depth=3) pada residual validasi	RF dilatih pada residual out-of-fold validasi temporal LSTM	Mengoreksi bias sistematis prediksi model sekuensial
Ensemble (Voting & Weighted Average)	Voting: rata-rata bobot sama; Weighted: bobot invers MAE_{val} (Pers. 2)	Bobot ditentukan murni dari error validasi temporal	Menstabilkan variansi prediksi lintas karakteristik SKU

3.5 Metrik Evaluasi Multi-SKU dan Simulasi Rekomendasi Persediaan

Mengingat nilai MAPE menjadi tidak terdefinisi ketika permintaan aktual bernilai nol (y_t = 0) dan mengalami lonjakan asimetris saat permintaan aktual bernilai sangat kecil, evaluasi model dilengkapi dengan metrik Weighted Absolute Percentage Error (WAPE) dan Mean Absolute Scaled Error (MASE) di samping MAE dan RMSE [3], [15], [26]:

$$MAE = (1/n) \Sigma |y_t - \hat{y}_t| , \qquad RMSE = \sqrt{[(1/n) \Sigma (y_t - \hat{y}_t)^2]} \quad (3)$$

$$WAPE = (\Sigma_\{t=1\}^\{n\} |y_t - \hat{y}_t| / \Sigma_\{t=1\}^\{n\} y_t) \times 100\% , \qquad MASE = MAE / [(1/(N_train - 1)) \Sigma_\{k=2\}^\{N_train\} |y_k - y_\{k-1\}|] \quad (4)$$

 ISSN:2302-8149
e-ISSN:2540-9719

Nilai MASE dapat dihitung ketika denominator absolut perubahan antar-langkah pada training series bernilai lebih besar dari nol. MASE menormalisasi MAE pengujian terhadap rata-rata kesalahan metode Naïve in-sample, sehingga metrik ini bebas skala (scale-free) dan dapat dibandingkan secara langsung antar-SKU [15]. Pada SKU dengan denominator pelatihan konstan (nol), pelaporan MASE ditiadakan secara eksplisit. Sementara itu, untuk metrik WAPE, persentase tidak dihitung apabila total permintaan aktual pengujian bernilai nol (pembagi nol); pada kasus tersebut, performa model dievaluasi murni menggunakan MAE dan RMSE untuk 19 SKU terdampak.

Hasil peramalan digunakan untuk menyusun simulasi parameter pengadaan berupa stok pengaman (Safety Stock / SS) dan titik pemesanan kembali (Reorder Point / ROP) sesuai Persamaan sesuai Persamaan (5) [1], [18]:

$$SS = Z \times \sigma_{e,val} \times \sqrt{L}\,, \qquad ROP = \hat{D}_{LT} + SS \tag{5}$$

Dalam Persamaan (5), Z = 1,65 merepresentasikan faktor kuantil untuk tingkat layanan siklus 95%, L = 2 minggu adalah asumsi lead time operasional, dan $\hat{D}_{LT}$ adalah akumulasi prediksi selama periode lead time. Hal yang paling krusial, estimasi ketidakpastian prediksi ($\sigma_{e,val}$) diperoleh murni dari simpangan baku kesalahan prediksi pada periode validasi pengembangan, bukan dari residual holdout test. Hal ini menjamin bahwa parameter persediaan dihitung menggunakan informasi yang benar-benar tersedia sebelum final test dievaluasi.

4 Hasil dan Pembahasan

4.1 Hasil Eksperimen

4.1.1 Distribusi Pola Permintaan 145 SKU (Matriks Syntetos-Boylan)

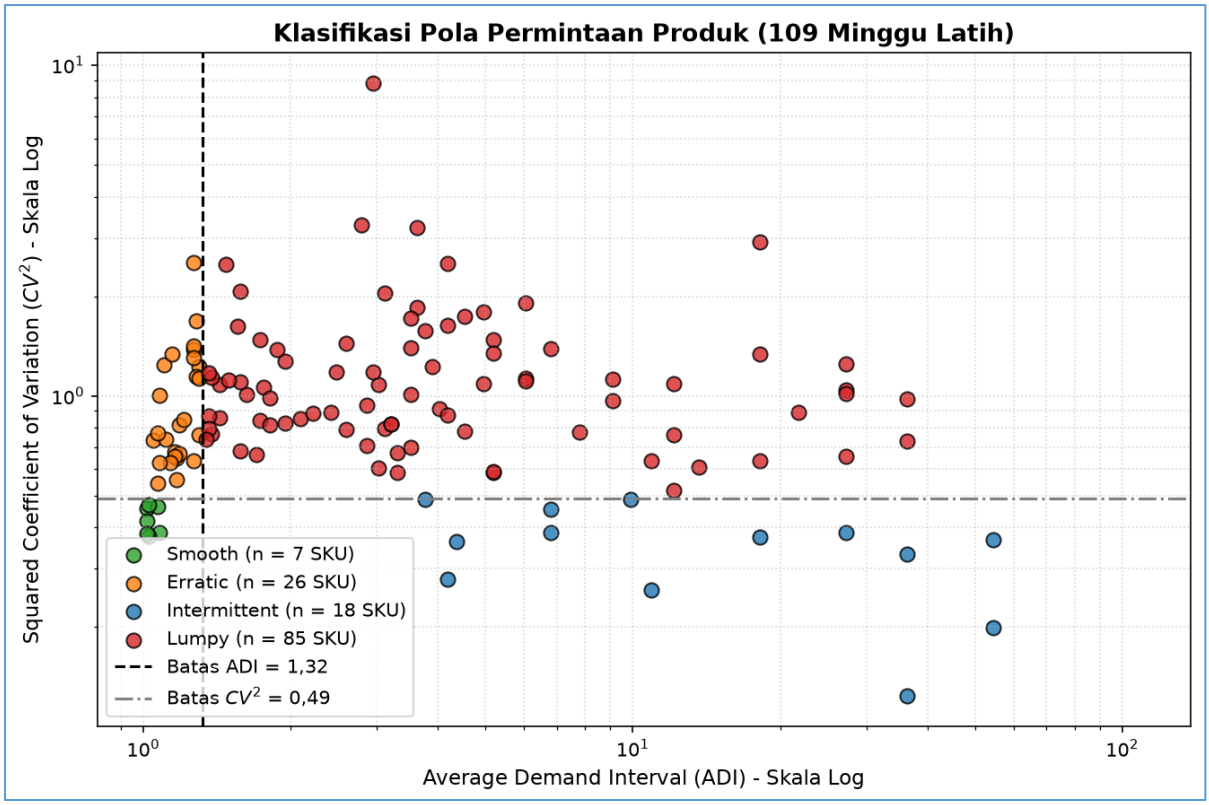
Pemetaan parameter ADI dan CV² terhadap 145 SKU pada 109 minggu data pelatihan menunjukkan struktur inventori yang didominasi oleh permintaan musiman/sporadis, sebagaimana disajikan pada Tabel 4 dan divisualisasikan dalam kuadran logaritmik pada Gambar 2.

Tabel 4 Rekonsiliasi kelayakan dan klasifikasi pola permintaan 145 SKU (periode pelatihan 109 minggu)

Pola Permintaan (Kriteria ADI & CV²)	Total SKU	MASE Valid (Denom > 0)	SKU Uji Positif (> 0)	SKU Uji Nol (= 0)	Total Baris Transaksi	Total Unit Keluar (139 Mgg)
Smooth (ADI ≤ 1,32; CV² ≤ 0,49)	7	7	7	0	4.868	2.493.797
Erratic (ADI ≤ 1,32; CV² > 0,49)	26	26	26	0	8.175	4.121.420
Intermittent (ADI > 1,32; CV² ≤ 0,49)	18	18	8	10	228	57.281
Lumpy (ADI > 1,32; CV² > 0,49)	85	85	76	9	6.337	3.557.466
Cold-Start / Inactive (Train = 0)	9	0	8	1	44	21.315
Total / Keseluruhan	**145**	**136**	**125**	**20**	**19.652**	**10.251.279**

Tabel 4 merangkum 145 SKU yang terbagi atas 136 SKU aktif (93,79%) dengan denominator MASE valid dan 9 SKU cold-start (6,21%) tanpa riwayat pelatihan yang dipisahkan dari pemodelan deret waktu. Dari sisi jumlah item, pola Lumpy mendominasi lebih dari separuh katalog (58,62%), diikuti Erratic (17,93%), Intermittent (12,41%), dan Smooth (4,83%). Sebaliknya, dari sisi volume fisik keluar, 33 SKU pada kategori Smooth dan Erratic (22,76% item) menyumbang 64,53% dari total 10.251.279 unit barang gudang. Untuk menghindari bias seleksi, sebanyak 19 SKU Intermittent (10 SKU) dan Lumpy (9 SKU) yang tidak memiliki transaksi selama 28 minggu pengujian tetap dipertahankan dalam evaluasi MAE dan MASE (tanpa perhitungan WAPE karena pembagi bernilai nol). Karakteristik 8 SKU representatif (dua produk bervolume tertinggi pada tiap kuadran) dirinci pada Tabel 5 dan pola deret waktunya divisualisasikan pada Gambar 3.

 ISSN:2302-8149
e-ISSN:2540-9719

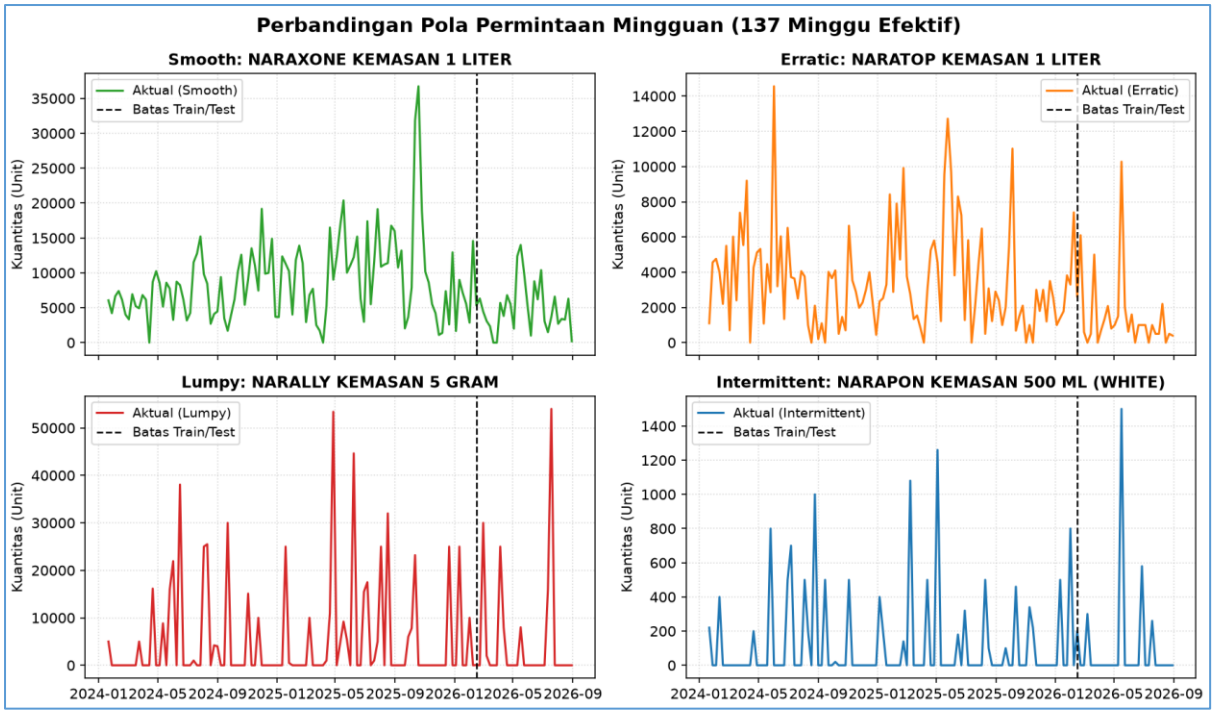


Gambar 2 Pemetaan klasifikasi pola permintaan produk pada matriks syntetos-boylan (109 minggu latih)

Tabel 5 Karakteristik statistik 8 SKU representatif berdasarkan volume tertinggi per kategori pola permintaan

Kategori Pola	Nama Produk (SKU Representatif)	ADI Latih	CV² Latih	Aktif Latih (/109)	Minggu Nol (%)	Aktif Uji (/28)	Total Volume (Unit)
Smooth	NARAXONE KEMASAN 1 LITER	1,02	0,42	107	1,83%	26	1.093.767
Smooth	NARA UP KEMASAN 1 LITER	1,02	0,46	107	1,83%	26	611.220
Erratic	NARATOP KEMASAN 1 LITER	1,07	0,54	102	6,42%	23	428.320
Erratic	NARA REL KEMASAN 100 ML	1,11	0,74	98	10,09%	21	342.600
Lumpy	NARALLY KEMASAN 5 GRAM	2,87	0,71	38	65,14%	7	726.560
Lumpy	NARA GENT KEMASAN 50 ML	1,36	0,80	80	26,61%	18	564.300
Intermittent	NARAPON KEMASAN 500 ML (WHITE)	3,76	0,49	29	73,39%	4	15.480
Intermittent	NARABAS 250 ML + SURFAKTAN 100 ML	4,19	0,28	26	76,15%	9	7.180

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 3 Perbandingan pola permintaan deret waktu mingguan pada 4 SKU representatif (137 minggu efektif)

4.1.2 Evaluasi Performa Model Lintas Kategori SKU, Ablasi Fitur, dan Horizon Prediksi

Untuk mencegah bias seleksi model (model selection bias), penetapan model terbaik pada masing-masing kategori pola permintaan dilakukan murni berdasarkan kinerja 5-fold rolling-origin cross-validation pada 109 minggu data pelatihan, bukan dari data pengujian [12], [26]. Tabel 6 menyajikan perbandingan median MASE seluruh sepuluh model pada tahap validasi temporal beserta model terpilih per kategori dan kinerjanya saat dievaluasi pada 28 minggu data uji holdout.

Tabel 6 Perbandingan median MASE data validasi (5-fold rolling-origin CV) dan kinerja model terpilih pada data uji (136 SKU aktif)

Kategori Pola	Naïve	MA-4	ETS	SBA	RF	SVR	LSTM	Hyb.	Vot.	Wght.	Terpilih (Val)	MASE Uji
Smooth (n=7)	1,033	0,841	0,825	0,801	0,849	0,855	0,817	0,856	0,801	**0,799**	**Weighted**	0,901
Erratic (n=26)	1,009	0,854	0,843	0,785	0,833	0,850	**0,779**	0,854	0,792	0,788	**LSTM**	0,797
Intermittent (n=18)	1,067	1,054	1,071	1,070	1,002	1,242	**0,993**	1,185	1,131	1,099	**LSTM**	0,633
Lumpy (n=85)	1,059	0,976	0,959	0,908	0,935	0,887	0,893	0,978	0,884	**0,878**	**Weighted**	0,825

Selanjutnya, Tabel 7 menyajikan perbandingan penuh median MASE seluruh sepuluh model pada periode pengujian holdout (28 minggu) per kategori pola permintaan, sekaligus evaluasi gabungan lintas 136 SKU (Median dan Mean MASE) guna menguji apakah strategi pemilihan model per kategori berbasis validasi (Pattern-Aware) mengungguli penerapan model tunggal secara seragam.

Tabel 7 Perbandingan median MASE periode pengujian (28 minggu holdout) per kategori pola dan evaluasi gabungan 136 SKU

Kategori / Evaluasi Gabungan	Naïve	MA-4	ETS	SBA	RF	SVR	LSTM	Hyb.	Vot.	Wght.	Pattern-Aware
Smooth (n=7)	1,000	0,845	**0,808**	0,810	0,922	0,935	0,968	1,087	0,901	0,901	0,901
Erratic (n=26)	0,790	0,732	0,751	0,745	0,846	**0,726**	0,797	0,924	0,790	0,789	0,797
Intermittent (n=18)*	**0,000***	0,085*	0,091*	0,826	0,643	1,388	0,633	0,750	0,837	0,759	0,633
Lumpy (n=85)	0,944	0,831	**0,804**	0,878	0,903	0,925	0,809	0,921	0,845	0,825	0,825

 ISSN:2302-8149
e-ISSN:2540-9719

Kategori / Evaluasi Gabungan	Naïve	MA-4	ETS	SBA	RF	SVR	LSTM	Hyb.	Vot.	Wght.	Pattern-Aware
Median Gabungan (136 SKU)	0,845	**0,778**	0,780	0,815	0,871	0,919	0,793	0,922	0,835	0,820	**0,804**
Mean Gabungan (136 SKU)	1,056	1,010	0,985	1,072	1,003	1,150	**0,976**	1,059	1,033	1,012	**0,992**

**Catatan: Sebanyak 10 dari 18 SKU pada kategori Intermittent (55,56%) tidak mencatat transaksi selama 28 minggu pengujian (Tabel 4), sehingga model jangka pendek yang memprediksi nol (Naïve, MA-4, ETS) memperoleh median MASE mendekati atau sama dengan nol.*

Hasil Tabel 6 dan Tabel 7 memperlihatkan temuan metodologis yang krusial. Berdasarkan data validasi (Tabel 6), Weighted Ensemble terpilih sebagai model terbaik pada pola Smooth (0,799) dan Lumpy (0,878), sementara LSTM terpilih pada pola Erratic (0,779) dan Intermittent (0,993). Namun, ketika diterapkan pada 28 minggu data uji (Tabel 7), terjadi pergeseran peringkat antar-model (validation-to-test shift): pada kategori Smooth dan Lumpy, ETS mencatat median MASE uji terendah (0,808 dan 0,804); pada kategori Erratic, SVR menjadi yang terbaik (0,726); sedangkan pada kategori Intermittent, Naïve (0,000), MA-4 (0,085), dan ETS (0,091) memperoleh nilai median sangat rendah akibat dominasi 10 SKU uji bernilai nol. Selain itu, kinerja Weighted Ensemble tidak selalu berada pada tiga besar di seluruh kuadran—meskipun menempati peringkat ketiga pada Lumpy (0,825), peringkatnya berada di posisi keempat pada Smooth (0,901), kelima pada Erratic (0,789), dan ketujuh pada Intermittent (0,759). Demikian pula, SVR tidak menunjukkan stabilitas merata karena meskipun unggul pada Erratic (0,726), kinerjanya relatif rendah pada Smooth (0,935), Lumpy (0,925), dan Intermittent (1,388). Pada evaluasi gabungan 136 SKU, strategi pemilihan model per kategori berbasis validasi (Pattern-Aware) mencatat median MASE 0,804 (Mean 0,992) yang melampaui 7 dari 10 model tunggal (termasuk Croston-SBA 0,815, Weighted Ensemble 0,820, Voting 0,835, Naïve 0,845, RF 0,871, SVR 0,919, dan Hybrid 0,922), tetapi masih sedikit berada di bawah tiga model tunggal teratas, yaitu MA-4 (0,778), ETS (0,780), dan LSTM (0,793; Mean terendah 0,976). Temuan ini membuktikan bahwa seleksi berbasis validasi tetap lebih adil secara metodologis dibanding seleksi post-hoc dari data uji, sekaligus menegaskan daya saing metode statistik sederhana pada deret waktu pendek.

Untuk mengevaluasi perilaku arsitektur secara mendalam, Tabel 8 menyajikan perbandingan lengkap pada SKU Smooth terbesar (NARAXONE KEMASAN 1 LITER) dengan horizon h = 1 minggu, beserta eksperimen ablasi fitur pada Tabel 9. Pada kalibrasi 5-fold CV di dalam 109 minggu data pelatihan, nilai kesalahan validasi temporal sebesar MAE_{val,RF} = 4.581,77 unit, MAE_{val,SVR} = 4.865,59 unit, dan MAE_{val,LSTM} = 5.329,47 unit, menghasilkan pembobot ensemble w_{RF} = 0,357, w_{SVR} = 0,336, dan w_{LSTM} = 0,307 sesuai Persamaan (2).

Tabel 8 Hasil evaluasi model pada produk utama (NARAXONE 1L) dengan validasi rolling-origin (horizon h = 1 minggu)

Kelompok & Nama Algoritma	MAE (Unit)	RMSE (Unit)	WAPE (%)	MASE	MAPE Non-Nol (%)
Naïve Forecasting (Baseline)	3.217,14	4.031,33	64,83%	0,71	194,89%
Moving Average 4 Minggu (Baseline)	3.217,86	3.948,27	64,85%	0,71	171,66%
Exponential Smoothing / ETS (Baseline)	3.101,13	3.718,23	62,50%	0,68	173,10%
Croston-SBA (Baseline Intermittent)	2.956,94	3.675,15	59,59%	0,65	147,24%
Support Vector Regression (SVR)	3.597,08	4.242,65	72,49%	0,79	160,73%
Random Forest (RF Recency-Weighted)	3.870,89	4.527,42	78,01%	0,85	200,46%
Long Short-Term Memory (LSTM)	2.932,66	3.517,59	59,10%	0,64	177,80%
Hybrid Model (LSTM + OOF Residual RF)	3.155,47	3.753,60	63,59%	0,69	211,81%
Voting Regressor (Simple Average Ensemble)	3.348,80	3.934,19	67,49%	0,74	177,60%

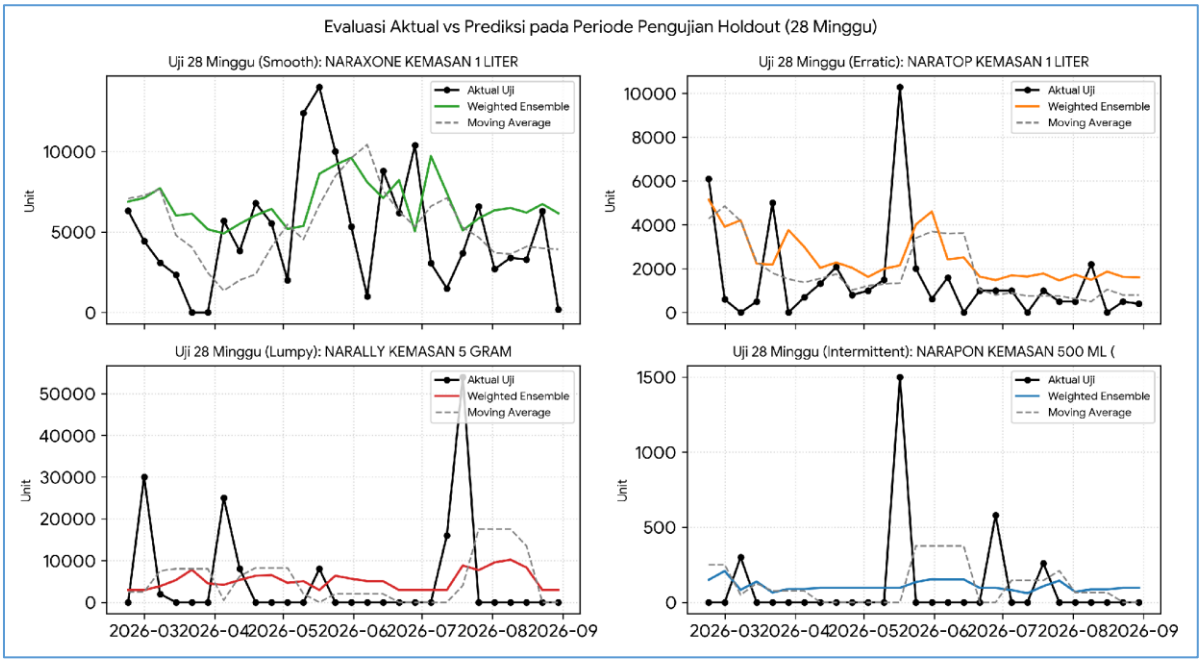
 ISSN:2302-8149
e-ISSN:2540-9719

Kelompok & Nama Algoritma	MAE (Unit)	RMSE (Unit)	WAPE (%)	MASE	MAPE Non-Nol (%)
Weighted Average Ensemble (Bobot Validasi)	3.378,81	3.962,67	68,09%	0,74	178,19%

Tabel 9 Hasil eksperimen ablasi fitur (E2) dan evaluasi horizon prediksi (h = 1, 2, dan 4 Minggu) pada 4 SKU representatif

Skenario Eksperimen Ablasi & Horizon	Mean MAE (Unit)	Mean MASE	Mean WAPE (%)	Temuan Evaluasi Teknis
Ablasi Fitur F1: Lag-Only (Lag 1 & Lag 2)	3.582,63	0,841	139,60%	Konteks memori terlalu pendek untuk menangkap tren bulanan
Ablasi Fitur F2: Lag + Rolling Stats (Lag 1,2,4 + RM4/RS4)	3.459,13	0,816	134,56%	Kombinasi fitur terbaik; meredam noise mingguan
Ablasi Fitur F3: Lag + Kalender (Bulan & Kuartal)	3.647,06	0,884	148,18%	Penambahan fitur kalender tidak memberikan peningkatan performa dari F2
Horizon Prediksi h = 1 Minggu (One-Step-Ahead)	3.459,13	0,816	134,56%	Akurasi tertinggi untuk pembaruan stok jangka sangat pendek
Horizon Prediksi h = 2 Minggu (Sesuai Lead Time L=2)	3.596,20	0,849	139,90%	Meningkatkan WAPE sebesar 5,34 poin persentase
Horizon Prediksi h = 4 Minggu (Proyeksi Bulanan)	3.840,55	0,906	149,41%	Akurasi menurun seiring hilangnya informasi transaksi terbaru

Perlu dicatat bahwa nilai rata-rata WAPE pada Tabel 9 melampaui 100% (134,56%–149,41%) karena rata-rata tersebut mencakup SKU berpola Intermittent dan Lumpy (NARAPON 500 ML dan NARALLY 5 GRAM) yang memiliki total permintaan uji kecil dan didominasi minggu bernilai nol. Pada karakteristik permintaan sporadis tersebut, akumulasi kesalahan prediksi positif kecil pada minggu tanpa transaksi (overforecasting moderat) ditambah deviasi besar saat terjadi lonjakan pesanan mendadak (underforecasting ekstrem) menyebabkan total kesalahan absolut melebihi total volume aktual selama 28 minggu pengujian (Gambar 4). Oleh karena itu, interpretasi akurasi lintas-SKU pada penelitian ini bertumpu pada metrik bebas skala MASE.



Gambar 4 Perbandingan nilai aktual terhadap prediksi model pada periode pengujian holdout (28 Minggu) untuk 4 SKU representatif

 ISSN:2302-8149
e-ISSN:2540-9719

4.1.3 Simulasi Ketersediaan dan Rekomendasi Persediaan (Safety Stock & ROP)

Untuk menguji dampak prediksi pada keputusan inventori (inventory simulation), dilakukan simulasi aliran ketersediaan barang selama 28 minggu pengujian menggunakan model Weighted Ensemble pada keempat SKU representatif. Tabel 10 menyajikan hasil verifikasi kalkulasi stok pengaman (Safety Stock) berdasarkan ketidakpastian prediksi validasi temporal (σ_{e,val}), titik pemesanan kembali (ROP), jumlah minggu mengalami kekosongan stok (stockout), serta indikator Cycle Service Level (CSL)—yakni proporsi minggu tanpa stockout selama periode pengujian—bukan unit fill rate yang mengukur proporsi pemenuhan unit permintaan [18]. Evaluasi perbandingan berbasis prediksi mingguan dinamis maupun rata-rata proyeksi menghasilkan jumlah minggu stockout dan CSL yang identik pada keempat SKU.

Tabel 10 Simulasi safety stock (validasi temporal σ_{e,val}), reorder point, dan cycle Service level (CSL) pada 28 Minggu Uji (L = 2 minggu, target SL = 95%)

Nama Produk (SKU Representatif)	Validasi σ_{e,val} (Unit)	Safety Stock (Unit)	Reorder Point (ROP)	Minggu Stockout	Cycle Service Level (CSL)*
NARAXONE KEMASAN 1 LITER (Smooth)	6.063,12	14.148,01	29.089,48	0 dari 28 mgg	100,00%
NARATOP KEMASAN 1 LITER (Erratic)	2.896,08	6.757,86	12.174,26	1 dari 28 mgg	96,43%
NARALLY KEMASAN 5 GRAM (Lumpy)	11.603,37	27.075,91	36.408,29	1 dari 28 mgg	96,43%
NARAPON KEMASAN 500 ML (Intermittent)	272,69	636,30	821,85	1 dari 28 mgg	96,43%

**Catatan: Cycle Service Level (CSL) dihitung sebagai proporsi minggu tanpa stockout dari total 28 minggu pengujian ((28 − jumlah minggu stockout) / 28 × 100%).*



Gambar 5 Tampilan dasbor antarmuka streamlit untuk klasifikasi multi-SKU serta simulasi safety stock dan reorder point

4.2 Pembahasan

Temuan eksperimen menegaskan pentingnya protokol seleksi model berbasis data validasi dan kehati-hatian dalam menggeneralisasi kinerja algoritma lintas pola permintaan. Sebagaimana ditunjukkan pada Tabel 6 dan Tabel 7, strategi pemilihan model per kategori berbasis validasi (Pattern-Aware) menghasilkan kinerja gabungan 136 SKU yang solid (median MASE 0,804; mean 0,992) dan mengungguli mayoritas model tunggal, namun metode statistik parsimonius seperti MA-4 (0,778) dan ETS (0,780) serta LSTM (0,793) tetap unggul tipis pada periode uji, sejalan dengan temuan empiris literatur peramalan mutakhir [19], [22]. Pergeseran peringkat dari validasi ke

 ISSN:2302-8149
e-ISSN:2540-9719

pengujian ini terjadi karena deret waktu pelatihan relatif pendek (109 minggu) dengan dominasi minggu bernilai nol pada kuadran Intermittent dan Lumpy. Di sisi simulasi inventori (Tabel 10), penerapan stok pengaman berbasis ketidakpastian validasi ($\sigma_{e,val}$) mampu menjaga Cycle Service Level (CSL) pada tingkat 100,00% untuk produk Smooth (NARAXONE 1L, 0 minggu stockout) dan 96,43% untuk ketiga SKU lainnya (NARATOP 1L, NARALLY 5 GRAM, dan NARAPON 500 ML, masing-masing 1 minggu stockout dari 28 minggu pengujian), melampaui target tingkat layanan siklus 95%.

Meskipun demikian, hasil simulasi Safety Stock, Reorder Point, dan CSL pada Tabel 10 memiliki keterbatasan metodologis penting. Simulasi dilakukan menggunakan skenario asumsi waktu tunggu statis (L = 2 minggu) tanpa ketersediaan data aktual stok fisik di tangan (on-hand inventory), jadwal kedatangan pasokan riil, maupun catatan antrean pesanan tertunda/hilang (backorders dan lost sales) dari distributor [11], [23]. Selain itu, metrik CSL hanya mengukur frekuensi minggu tanpa kehabisan stok, bukan besaran unit permintaan yang gagal dipenuhi saat stockout terjadi. Oleh sebab itu, keluaran dasbor Streamlit (Gambar 5) berfungsi sebagai alat bantu keputusan indikatif bagi perencanaan pengadaan dan belum dapat diinterpretasikan secara definitif sebagai pembuktian penurunan biaya persediaan aktual di lapangan.

Dari perspektif implikasi manajerial, fakta bahwa 85 SKU (58,62% katalog) termasuk kategori Lumpy menuntut kebijakan persediaan terdiferensiasi dan tidak dapat disamakan dengan produk berpola Smooth. Pada produk Lumpy bervolume tinggi seperti NARALLY KEMASAN 5 GRAM (65,14% minggu nol dengan lonjakan ekstrem hingga di atas 50.000 unit), tingginya simpangan baku kesalahan validasi ($\sigma_{e,val}$ = 11.603,37 unit) mendorong kebutuhan Safety Stock sebesar 27.075,91 unit dan Reorder Point 36.408,29 unit agar risiko stockout pada puncak musim tanam tetap terkendali. Apabila kebijakan penyangga stok besar diterapkan seragam pada seluruh 85 SKU Lumpy, gudang akan terbebani modal kerja tertahan, kapasitas ruang simpan, dan risiko kedaluwarsa produk pestisida. Oleh karena itu, manajemen disarankan mengombinasikan hasil klasifikasi ADI-CV^2 dengan segmentasi nilai (analisis ABC-XYZ) [1], [18], [23]: menerapkan buffer Safety Stock terukur berbasis $\sigma_{e,val}$ dan pemantauan ROP berkala khusus untuk SKU Lumpy bernilai kontribusi tinggi, menerapkan kebijakan pemesanan berbasis pesanan (make-to-order / order-on-demand) atau kesepakatan pengiriman cepat dengan pemasok bagi SKU Lumpy bernilai rendah, serta memperpendek siklus evaluasi ulang ROP menjelang periode puncak distribusi musiman guna menekan risiko stockout tanpa memicu overstock berkepanjangan.

5 Kesimpulan

Berdasarkan evaluasi komparatif sepuluh metode peramalan permintaan pada data transaksi PT Agro Sarana Utama (19.652 rekaman bersih mencakup 145 SKU selama 137 minggu efektif), disimpulkan bahwa pemetaan pola permintaan berbasis matriks Syntetos-Boylan pada data pelatihan merupakan tahapan fundamental. Struktur katalog didominasi oleh pola Lumpy (58,62% item), sementara 64,53% total volume fisik keluar justru berasal dari 33 SKU berpola Smooth dan Erratic. Evaluasi yang memisahkan 109 minggu latih (5-fold rolling-origin cross-validation) dan 28 minggu holdout test tanpa bias seleksi menunjukkan bahwa seleksi model berbasis validasi memilih Weighted Ensemble pada pola Smooth (0,799) dan Lumpy (0,878) serta LSTM pada Erratic (0,779) dan Intermittent (0,993). Pada data pengujian, model terpilih validasi tersebut menghasilkan median MASE 0,901 (Smooth), 0,797 (Erratic), 0,633 (Intermittent), dan 0,825 (Lumpy), dengan evaluasi gabungan 136 SKU sebesar 0,804 (Mean 0,992) yang mengungguli tujuh model tunggal lainnya meskipun sedikit di bawah MA-4 (0,778), ETS (0,780), dan LSTM (0,793; Mean 0,976). Sementara itu, kinerja SVR unggul pada Erratic (0,726) tetapi kurang stabil pada Smooth (0,935), Lumpy (0,925), dan Intermittent (1,388), sedangkan Weighted Ensemble konsisten berada di papan atas pada Lumpy (0,825) namun menurun ke peringkat kelima pada Erratic (0,789). Integrasi hasil peramalan ke dalam simulasi Safety Stock dan Reorder Point berbasis ketidakpastian validasi temporal ($\sigma_{e,val}$) menghasilkan estimasi Cycle Service Level (CSL) sebesar 100,00% pada NARAXONE 1L dan 96,43% pada NARATOP 1L, NARALLY 5 GRAM, serta NARAPON 500 ML melalui dasbor Streamlit sebagai masukan perencanaan pengadaan gudang.

 ISSN:2302-8149
e-ISSN:2540-9719

Sebagai saran pengembangan untuk penelitian lanjutan, direkomendasikan tiga agenda utama. Pertama, melaksanakan pengujian model secara langsung di lapangan (live field testing) pada operasional gudang untuk memvalidasi dampak aktual rekomendasi peramalan terhadap penurunan stockout dan efisiensi biaya simpan. Kedua, mengintegrasikan variabel eksternal relevan sektor agroindustri, seperti data cuaca, curah hujan, siklus musim tanam, dan kalender aktivitas pertanian, guna meningkatkan akurasi prediksi lonjakan permintaan musiman. Ketiga, mengembangkan kerangka peramalan yang terintegrasi secara langsung dengan strategi pengendalian persediaan dinamis—seperti optimasi kebijakan (s, S) dan pemodelan biaya simpan maupun kekosongan—berbasis data aktual on-hand inventory dan catatan backorders.

Referensi

[1] S. Chopra and P. Meindl, *Supply Chain Management: Strategy, Planning, and Operation*, 6th ed. Boston, MA, USA: Pearson, 2016.

[2] S. Alam Khan and A. Erdoğdu, "*A Study on the Impact of Artificial Intelligence on Demand Forecasting in Food Industries*," *Journal of Life Economics*, Vol. 12, No. 1, pp. 45–58, 2025, DOI: 10.15637/jlecon.2682.

[3] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Melbourne, Australia: OTexts, 2021.

[4] A. A. Syntetos and J. E. Boylan, "*The Accuracy of Intermittent Demand Estimates*," *International Journal of Forecasting*, Vol. 21, No. 2, pp. 303–314, 2005, DOI: 10.1016/j.ijforecast.2004.10.001.

[5] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "*M5 Accuracy Competition: Results, Findings, and Conclusions*," *International Journal of Forecasting*, Vol. 38, No. 4, pp. 1346–1364, 2022, DOI: 10.1016/j.ijforecast.2021.11.013.

[6] R. Carbonneau, K. Laframboise, and R. Vahidov, "*Application of Machine Learning Techniques for Supply Chain Demand Forecasting*," *European Journal of Operational Research*, Vol. 184, No. 3, pp. 1140–1154, 2008, DOI: 10.1016/j.ejor.2006.12.004.

[7] L. Breiman, "Random Forests," *Machine Learning*, Vol. 45, No. 1, pp. 5–32, 2001, DOI: 10.1023/A:1010933404324.

[8] J. E. Boylan and A. A. Syntetos, *Intermittent Demand Forecasting: Context, Methods and Applications*. Chichester, UK: John Wiley & Sons, 2021.

[9] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780, 1997, DOI: 10.1162/neco.1997.9.8.1735.

[10] A. A. Levis and L. G. Papageorgiou, "*Customer Demand Forecasting via Support Vector Regression Analysis*," *Journal of Intelligent Manufacturing*, Vol. 16, No. 4–5, pp. 409–419, 2005, DOI: 10.1007/s10845-005-6892-1.

[11] G. D. Palgunadi, R. Fitriana, A. N. Habyba, and Y. Liang, "*Optimizing Demand Forecasting Method with Support Vector Regression for Improved Inventory Planning*," *Jurnal Optimasi Sistem Industri*, Vol. 24, No. 1, 2025.

[12] C. Bergmeir and J. M. Benítez, "*On the Use of Cross-Validation for Time Series Predictor Evaluation*," *Information Sciences*, Vol. 191, pp. 192–213, 2012, DOI: 10.1016/j.ins.2011.12.028.

[13] L. Cao and F. E. H. Tay, "*Support Vector Machines Experts for Time Series Forecasting*," *Neurocomputing*, Vol. 51, pp. 321–339, 2003, DOI: 10.1016/S0925-2312(02)00577-5.

[14] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "*Support Vector Regression Machines*," in *Advances in Neural Information Processing Systems*, Vol. 9, pp. 155–161, 1997.

[15] R. J. Hyndman and A. B. Koehler, "*Another Look at Measures of Forecast Accuracy*," *International Journal of Forecasting*, Vol. 22, No. 4, pp. 679–688, 2006, DOI: 10.1016/j.ijforecast.2006.03.001.

[16] J. D. Croston, "*Forecasting and Stock Control for Intermittent Dem*ands," *Operational Research Quarterly*, Vol. 23, No. 3, pp. 289–303, 1972.

[17] A. A. Syntetos, J. E. Boylan, and J. D. Croston, "*On the Categorization of Demand Patterns*," *Journal of the Operational Research Society*, Vol. 56, No. 5, pp. 495–503, 2005.

[18] E. A. Silver, D. F. Pyke, and D. J. Thomas, *Inventory and Production Management in Supply Chains*, 4th ed. Boca Raton, FL, USA: CRC Press, 2017.
[19] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "*Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward*," PLOS ONE, Vol. 13, No. 3, e0194889, 2018.
[20] S. Smyl, "*A Hybrid Method of Exponential Smoothing and Recurrent Neural Networks for Time Series Forecasting*," *International Journal of Forecasting*, Vol. 36, No. 1, pp. 75–85, 2020.
[21] H. Hewamalage, C. Bergmeir, and K. Bandara, "*Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions*," *International Journal of Forecasting*, Vol. 37, No. 1, pp. 388–427, 2021.
[22] F. Petropoulos et al., "*Forecasting: Theory and Practice*," *International Journal of Forecasting*, Vol. 38, No. 3, pp. 705–871, 2022.
[23] R. Fildes, S. Ma, and S. Kolassa, "*Retail Forecasting: Research and Practice*," *International Journal of Forecasting*, Vol. 38, No. 4, pp. 1283–1318, 2022.
[24] T. Januschowski, Y. Wang, K. Torkkola, T. Erkkilä, H. Hasson, and J. Gasthaus, "*Forecasting with Trees*," *International Journal of Forecasting*, Vol. 38, No. 4, pp. 1473–1481, 2022.
[25] X. Wang, R. J. Hyndman, F. Li, and Y. Kang, "*Forecast Combinations: An Over 50-Year Review*," *International Journal of Forecasting*, Vol. 39, No. 4, pp. 1518–1547, 2023.
[26] H. Hewamalage, K. Ackermann, and C. Bergmeir, "*Forecast Evaluation for Data Scientists: Common Pitfalls and Best Practices*," *Data Mining and Knowledge Discovery*, Vol. 37, No. 2, pp. 788–832, 2023.