Explainable Hierarchical Retail Demand Forecasting: Comparison Of Statistical And Machine Learning Methods

Abdul Rohim, Departement of Computer Science, STMIK Al Muslim, Bekasi,Jawa Barat
abdul.rohim.pranata18@gmail.com[1] *
Novi Wulandari, Department, Department of Informatics Management, STMIK Al Muslim, Bekasi,Jawa Barat wulandari.novi1@gmail.com[2]
Corresponding Author[*]: abdul.rohim.pranata18@gmail.com

Abstract

Background : Retail demand forecasting across multiple aggregation levels is essential for consistent strategic and operational decision-making.

Purpose : This study develops an explainable hierarchical forecasting framework for weekly retail sales using Seasonal Naive, XGBoost, and LightGBM base models with Base Forecast, Bottom-Up, MinT-OLS, and MinT-Shrink reconciliation.

Methodology : Unlike single-level retail studies, the framework jointly evaluates accuracy, cross-level coherence, and TreeSHAP interpretability at total, category, and subcategory levels under leakage-safe temporal validation. The Kaggle Superstore Sales Analysis dataset was aggregated into 209 weekly periods containing one total-sales series, three category series, and 17 subcategory series.

Findings : The final four weeks were reserved as a strict holdout, while the preceding 205 weeks were evaluated using five expanding-window rolling-origin folds with a four-week horizon. LightGBM with MinT-OLS was selected by cross-validation, achieving an MAE of 1,344.3577, RMSE of 2,465.7518, WAPE of 48.4886%, MASE of 2.5970, and zero coherence error. When refitted on the full development period, the selected configuration achieved a holdout WAPE of 63.9589%. TreeSHAP identified rolling_mean_13, lag_52, rolling_mean_8, rolling_std_13, and lag_26 as the strongest predictors.

Implications : The dependence plot revealed a nonlinear increase in the contribution of rolling_mean_13, moderated by annual-lag conditions.

Originality : These results demonstrate that reconciliation can improve volume-weighted accuracy and eliminate contradictory forecasts, although intermittent subcategory demand and holdout variability remain important limitations.

Keywords: *hierarchical forecasting, LightGBM, XGBoost, MinT, TreeSHAP*

INTRODUCTION

Demand forecasting is a critical component of retail business management because inventory, distribution, procurement, and resource allocation decisions depend on the company's ability to meet customer needs in the following period. Forecast errors can lead to overstock when inventory exceeds demand and stockouts when inventory availability is insufficient. This problem becomes more complex when a company manages multiple categories, subcategories, marketing regions, and sales patterns that change over time.

In practice, retail demand is not simply defined as a single aggregate time series. Management requires total sales forecasts for strategic planning, while operational units require forecasts for categories and subcategories to determine stock allocation. Therefore, a forecasting system must produce accurate and consistent predictions across all hierarchical levels.

This problem is known as hierarchical forecasting. In this approach, total sales are the sum of sales at lower levels. A prediction is considered coherent if the aggregate-level prediction value equals

the sum of its derivative predictions. Wickramasuriya et al. introduced Minimum Trace, or MinT, to reconcile predictions so that hierarchical consistency is maintained (Wickramasuriya et al., 2019). A review by Athanasopoulos et al. Studies have shown that forecast reconciliation is a crucial part of modern forecasting because organizational decisions are typically made at multiple levels of aggregation (Athanasopoulos et al., 2024).

Various approaches have been used for sales prediction, ranging from statistical methods to machine learning. XGBoost developed an efficient and scalable tree boosting system (Chen & Guestrin, 2016), while LightGBM introduced gradient boosting optimization through Gradient-based One-Side Sampling and Exclusive Feature Bundling (Ke et al., 2017). Comparative studies have shown that more complex models do not always automatically produce the best results; method selection must consider data characteristics and evaluation plans (Makridakis et al., 2018).

In the Indonesian context, sales forecasting research commonly compares SARIMA, Exponential Smoothing, Prophet, XGBoost, RNN, LSTM, GRU, and LightGBM (Arunika & Dewi, 2025; Putra & Hendry, 2022; Sunendar & Rianto, 2025; Suryawan et al., 2024; Hidayat et al., 2025; Amiri & Wardati, 2025; Hasanah et al., 2025; Nailendra et al., 2025; Permana, 2024; Gymnastiar & Anwar, 2025). These studies show that performance depends on seasonality, data granularity, feature engineering, and validation strategy. However, most evaluate one aggregate target or report model accuracy independently at each level. Consequently, two issues remain underexplored in Indonesian retail forecasting: whether predictions at total, category, and subcategory levels are mathematically coherent, and whether the drivers of machine-learning forecasts can be explained for managerial use. Algorithm comparison alone cannot resolve contradictory forecasts across organizational levels or show why a model changes its demand estimate.

This research addresses that gap through an explainable hierarchical retail-demand forecasting framework that compares Seasonal Naive, XGBoost, and LightGBM and evaluates Base Forecast, Bottom-Up, MinT-OLS, and MinT-Shrink reconciliation. SHAP is then used to explain the temporal signals underlying the strongest machine-learning model. The novelty is not a new forecasting algorithm; it is the integrated and chronology-safe evaluation of predictive accuracy, hierarchical coherence, and feature-level interpretability across three retail aggregation levels in one reproducible experimental design. This design provides evidence about both the statistical value of reconciliation and the operational meaning of the reconciled forecasts.

METHODOLOGY

1. Research Design

This study uses a quantitative experimental approach. Figure 1 summarizes nine sequential stages: (1) auditing transaction fields, missing values, duplicates, date coverage, and category consistency; (2) aggregating transaction-level sales into complete weekly series and assigning zero to weeks without sales; (3) constructing the total-category-subcategory hierarchy and its summing matrix; (4) generating lag, rolling-statistic, and calendar features using only information available before each forecast origin; (5) defining rolling-origin folds and a final holdout in chronological order; (6) fitting Seasonal Naive, XGBoost, and LightGBM base models independently for each series; (7) reconciling base forecasts using Bottom-Up, MinT-OLS, or MinT-Shrink; (8) evaluating MAE, RMSE, WAPE, MASE, and coherence error at every hierarchical level; and (9) applying SHAP to the selected XGBoost model to interpret global feature effects. This order prevents information from future weeks from entering feature construction or training.

Corresponding Authors : abdul.rohim.pranata18@gmail.com
Inkofar 2026 ©

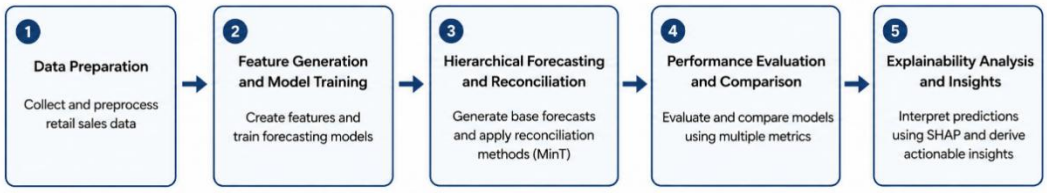


Figure 1. Explainable hierarchical retail demand forecasting research flow

2. Dataset

The dataset was obtained from the Superstore Sales Analysis collection published by Jain (2026) on Kaggle (https://www.kaggle.com/datasets/divyjain28/superstore-sales). The CSV file contains 10,194 transaction rows and 21 attributes, including order date, customer, product, category, subcategory, region, shipping method, sales, quantity, discount, and profit. It comprises 5,111 unique orders, 804 customers, 1,862 products, three categories, and 17 subcategories, with order dates from January 3, 2023, to December 30, 2026. Because Sample Superstore is a synthetic instructional benchmark, its dates are treated as ordered simulation timestamps rather than future real-world observations. Chronological splitting ensures that each forecast uses only information available before its origin; nevertheless, the synthetic data limit the external validity of the findings.

Table 1. Research dataset profile

Characteristics	Score
Transaction Line	10.194
Attributes	21
Unique Orders	5.111
Unique Customers	804
Unique Products	1.862
Categories	3
Subcategories	17
Regions	4
Weekly Period	209
Total Sales	2.326.534,35
Total Profit	292.296,81
Empty Values	0

Table 2. Sales distribution by category

Category	Total Sales	Total profit	Quantity	Transactions
Furniture	754.747,76	19.730,00	8.369	2.201
Office Supplies	731.893,31	126.023,44	23.268	6.128
Technology	839.893,28	146.543,38	7.017	1.865

3. Forecasting Target and Hierarchical Structure

The primary target of this study is weekly sales. The sales value for group g in week t is calculated as the sum of all transactions included in that group and period.

$$y_{g,t} = \Sigma_{i \in I(g,t)} \text{Sales}_i \quad (1)$$

Weekly aggregation was chosen because it balances temporal granularity and historical length. Daily data are highly volatile, whereas monthly aggregation would yield only 48 observations. The hierarchy contains total, category, and subcategory levels, producing 21 series (1 + 3 + 17). For each series, lag

Corresponding Authors : abdul.rohim.pranata18@gmail.com

features were generated at 1, 2, 4, 8, 13, 26, and 52 weeks. Rolling means and population-standard-deviation features were calculated over trailing 4-, 8-, and 13-week windows using only observations preceding the predicted week. Calendar variables comprised year, quarter, cyclical week sine-cosine terms, and cyclical month sine-cosine terms. One-hot series identifiers allowed each pooled tree model to distinguish aggregate and lower-level patterns. The first 52 observations of each series were excluded from supervised model fitting because they lacked complete annual-lag history.

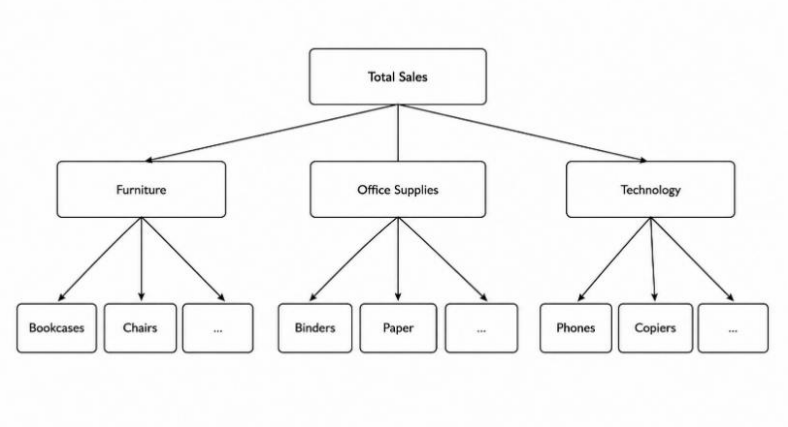


Figure 2. Retail sales hierarchy structure

4. Model and Reconciliation

Table 3. Reconciliation models and scenarios

Component	Functions in the experiment
Seasonal Naïve	Simple baseline based on historical seasonal values
XGBoost	Gradient boosting to learn non-linear relationships between temporal features
LightGBM	Efficient gradient boosting for XGBoost benchmarks
Base Forecast	Prediction of each series without reconciliation
Bottom-up	Aggregation of subcategory predictions into categories and totals
MinT-OLS	Coherent reconciliation based on ordinary least squares
MinT-Shrink	Coherent reconciliation with covariance shrinkage estimation

No automated hyperparameter search was performed; fixed settings were retained across all folds and series. XGBoost 3.1.2 used 250 estimators, maximum depth 5, learning rate 0.04, row subsampling 0.90, feature subsampling 0.90, squared-error objective, and random seed 42. LightGBM 4.6.0 used 250 estimators, 31 leaves, unrestricted maximum depth, learning rate 0.04, row subsampling 0.90 with subsample frequency 1, feature subsampling 0.90, and random seed 42. Both models used one processing thread to support deterministic replication

Base Forecast was retained as the unreconciled accuracy and coherence benchmark. Bottom-Up aggregates the 17 bottom-series predictions. MinT-OLS uses an identity forecast-error covariance matrix, whereas MinT-Shrink uses a Ledoit-Wolf shrinkage estimate of the in-sample one-step residual covariance matrix (Ledoit & Wolf, 2004). Both MinT variants use forecasts from every level through

Corresponding Authors : abdul.rohim.pranata18@gmail.com

the trace-minimizing projection. Top-Down and Middle-Out were excluded because they anchor the hierarchy at selected levels and require allocation proportions; ERM was excluded to avoid introducing another learned reconciliation component. These exclusions define the study scope and do not imply that the omitted methods are inferior.

Let $\hat{y}_t$ denote the 21-dimensional base-forecast vector, S the 21 × 17 summing matrix, and $\hat{b}_t$ the bottom-level forecast vector. A reconciled forecast is coherent when $\tilde{y}_t = S\,\hat{b}_t$. Let A contain the four aggregate nodes (one total and three categories), C(a) denote the bottom descendants represented by S for aggregate a, and H denote the forecast horizon. Coherence error is $CE = (1/H)\sum_t[(1/|A|)\sum_a|\tilde{y}_{a,t} - \sum_{j\in C(a)}\tilde{y}_{j,t}|]$. Only aggregate discrepancies enter CE; therefore, CE = 0 indicates exact cross-sectional coherence. MAE, RMSE, and WAPE were calculated from all observations within each reported level. MASE was first computed per series using the 52-week seasonal-naive training scale and then averaged, preventing values from different series from being concatenated into an invalid seasonal sequence.

RESULT AND DISCUSSION

1. Cross-Validation Results

The final four of the 209 weekly periods (December 13, 2026-January 3, 2027 after Sunday aggregation) were reserved as a strict holdout and were never used for model selection. The preceding 205 periods formed the development set. Five non-overlapping four-week test windows were evaluated with expanding training windows: Fold 1 trained on weeks 1-185 and tested on weeks 186-189; Fold 2 trained on weeks 1-189 and tested on weeks 190-193; Fold 3 trained on weeks 1-193 and tested on weeks 194-197; Fold 4 trained on weeks 1-197 and tested on weeks 198-201; and Fold 5 trained on weeks 1-201 and tested on weeks 202-205. Feature construction and covariance estimation were repeated within each fold using training data only. Models were ranked primarily by mean All-level WAPE across folds, with MAE, RMSE, per-series MASE, and coherence error as complementary measures. The selected configuration was then refitted on all 205 development weeks and evaluated once on weeks 206-209.

Table 4. Cross-validation results on the entire hierarchical series

No.	Model	Reconciliation	MAE	RMSE	WAPE(%)	MASE	CE
1	LightGBM	MinT-OLS	1,344.3577	2,465.7518	48.4886	2.5970	0.0000
2	LightGBM	MinT-Shrink	1,372.8948	2,507.9402	49.3952	2.4364	0.0000
3	XGBoost	MinT-OLS	1,381.1352	2,556.3540	49.6641	2.5971	0.0000
4	XGBoost	Base	1,407.6656	2,606.8171	50.5724	2.4336	2,360.1492
5	LightGBM	Base	1,403.4449	2,515.6138	50.7938	2.4475	2,655.6654
6	XGBoost	Bottom-Up	1,409.5590	2,509.1984	51.2419	2.4474	0.0000
7	XGBoost	MinT-Shrink	1,475.0365	2,799.5932	52.3965	2.2963	0.0000
8	LightGBM	Bottom-Up	1,472.3669	2,582.6271	54.4069	2.4702	0.0000
9	Seasonal Naive	MinT-Shrink	1,745.1494	3,485.1785	61.9506	2.1905	0.0000
10	Seasonal Naive	Base	1,745.1494	3,485.1785	61.9506	2.1905	0.0000
11	Seasonal Naive	Bottom-Up	1,745.1494	3,485.1785	61.9506	2.1905	0.0000
12	Seasonal Naive	MinT-OLS	1,745.1494	3,485.1785	61.9506	2.1905	0.0000

LightGBM with MinT-OLS achieved the lowest mean cross-validation WAPE (48.4886%), with an MAE of 1,344.3577, RMSE of 2,465.7518, MASE of 2.5970, and coherence error numerically equal to zero. XGBoost MinT-OLS ranked third with a WAPE of 49.6641%. The advantage of LightGBM was modest, suggesting that both boosted-tree models captured similar nonlinear temporal structure. MinT-Shrink did not improve LightGBM relative to MinT-OLS (49.3952% versus 48.4886% WAPE), indicating that the estimated residual covariance added variance without sufficient accuracy gain in this short hierarchy. MASE above 1 shows that, when every series is normalized by its own seasonal scale, errors remain substantial—especially for intermittent subcategories—even though volume-weighted WAPE improves.

Corresponding Authors : abdul.rohim.pranata18@gmail.com

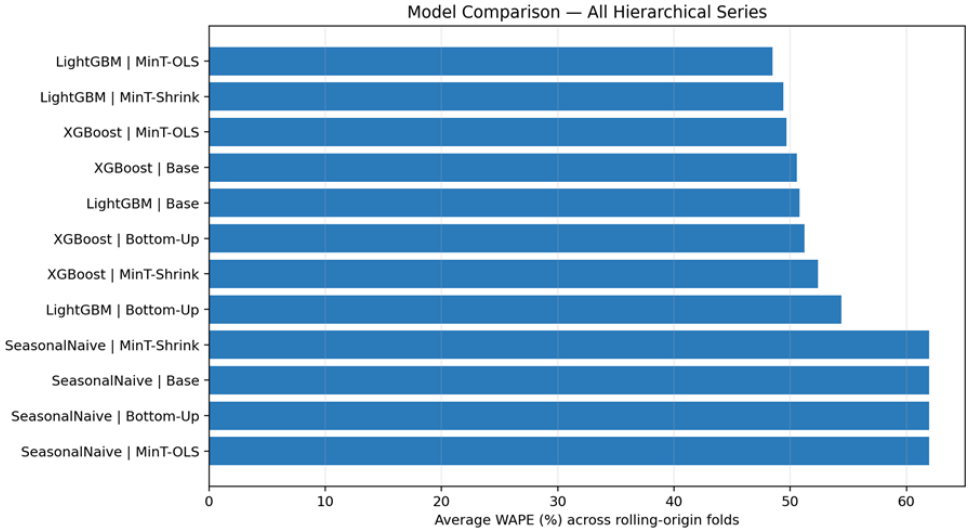


Figure 3. WAPE comparison of all models in the hierarchical series

Relative to LightGBM Base, MinT-OLS reduced WAPE from 50.7938% to 48.4886%, an absolute improvement of 2.3052 percentage points or 4.54% relatively. Relative to Seasonal Naive (61.9506%), the WAPE reduction was 21.73%. These gains show that reconciliation contributes beyond the LightGBM base learner, but the differing behavior of WAPE and MASE demonstrates that model assessment should not rely on one metric alone.

2. Evaluation at Each Level of the Hierarchy

Table 5. XGBoost performance at each level of the hierarchy

Reconciliation	Level	MAE	RMSE	WAPE(%)	MASE
Base	Total	5,919.3409	7,210.7684	28.5054	1.1700
Base	Category	2,690.6156	3,655.4395	43.1680	1.1678
Base	Sub category	910.6562	1,495.0723	80.7081	2.7484
Base	All	1,403.4449	2,515.6138	50.7938	2.4475
MinT-OLS	Total	5,605.3866	6,916.0120	27.4815	1.1090
MinT-OLS	Category	2,683.3913	3,667.2306	42.7884	1.1648
MinT-OLS	Sub category	857.4089	1,473.5241	75.1958	2.9373
MinT-OLS	All	1,344.3577	2,465.7518	48.4886	2.5970

Applying MinT-OLS improved LightGBM WAPE at every level: from 28.5054% to 27.4815% at Total, from 43.1680% to 42.7884% at Category, from 80.7081% to 75.1958% at Subcategory, and from 50.7938% to 48.4886% overall. The largest absolute improvement occurred at the intermittent subcategory level; nevertheless, a WAPE of 75.1958% and MASE of 2.9373 are too high for automatic replenishment. These forecasts should support exception-based review, safety-stock decisions, and managerial judgment rather than directly trigger purchase orders. Aggregate forecasts are more suitable for capacity, budget, and broad allocation planning.

Table 6. Subcategory with the highest proportion of no-sales weeks

Sub category	Week without sales (%)
Technology - Copiers	72,73
Technology - Machines	61,24

Corresponding Authors : abdul.rohim.pranata18@gmail.com

Office Supplies - Supplies	41,15
Office Supplies - Envelopes	36,36
Office Supplies - Fasteners	36,36
Furniture - Bookcases	34,93
Furniture - Tables	28,23

The average proportion of non-transaction weeks across 17 subcategories reached 23.73%. This explains part of the high subcategory WAPE because small nominal errors become large percentage errors when actual sales approach zero. Croston, SBA, and TSB are established intermittent-demand benchmarks; they were not included in the present model comparison, so the current results do not establish superiority over specialized intermittent-demand methods. Future work should compare reconciled machine-learning forecasts directly with these methods, particularly for Copiers, Machines, Supplies, Envelopes, and Fasteners.

5. Prediction Coherence

Table 7. Impact of reconciliation on coherence error

Reconciliation	Coherence Error	
	XGBoost	LightBM
Base	2,360.1492	2,655.6654
Bottom-up	0,0000	0,0000
MinT-OLS	0,0000	0,0000
MnT-Shrink	0,0000	0,0000

The corrected aggregate-node coherence error was 2,360.1490 for XGBoost Base and 2,655.6654 for LightGBM Base. Bottom-Up, MinT-OLS, and MinT-Shrink reduced the error to numerical zero. LightGBM MinT-OLS combined this coherence with the lowest cross-validation WAPE. This agrees with Wickramasuriya et al. (2019) and Athanasopoulos et al. (2024), who emphasize that reconciliation uses information across levels while enforcing aggregation constraints. The result extends those findings to pooled boosted-tree retail forecasts, although it is limited to one synthetic dataset and a small hierarchy.

6. Final Holdout Set

Table 9. Final holdout results for the entire series

No.	Model	Reconciliation	MAE	RMSE	WAPE(%)	MASE	CE
1	XGBoost	MinT-OLS	1,233.8613	1,798.8662	53.7893	1.6189	0.0000
2	XGBoost	MinT-Shrink	1,288.8642	2,236.9799	56.1871	1.2088	0.0000
3	XGBoost	Base	1,317.2909	2,458.1104	57.4263	1.1912	0.0000
4	LightGBM	MinT-Shrink	1,317.2909	2,458.1104	57.4263	1.1912	0.0000
5	LightGBM	MinT-OLS	1,317.2909	2,458.1104	57.4263	1.1912	0.0000
6	LightGBM	Base	1,317.2909	2,458.1104	57.4263	1.1912	0.0000
7	LightGBM	Bottom-Up	1,467.1386	2,444.2630	63.9588	1.4454	0.0000
8	XGBoost	Bottom-Up	1,506.7308	2,536.8790	65.6848	1.3567	2,406.9453
9	Seasonal Naive	Base	1,602.0517	2,476.1959	69.8403	1.5506	3,534.5558
10	Seasonal Naive	Bottom-Up	1,660.1994	2,811.2168	72.3752	1.3763	0.0000
11	Seasonal Naive	MinT-Shrink	1,670.1880	2,824.5209	72.8106	1.4481	0.0000
12	Seasonal Naive	MinT-OLS	1,920.8308	3,508.2896	83.7372	1.5948	0.0000

The configuration selected before inspecting the holdout—LightGBM MinT-OLS—achieved an MAE of 1,467.1386, RMSE of 2,444.2630, WAPE of 63.9589%, MASE of 1.4454, and zero coherence error. It improved upon LightGBM Base WAPE (65.6848%) but was outperformed on this four-week period

Corresponding Authors : abdul.rohim.pranata18@gmail.com

by Seasonal Naive (57.4263%) and, ex post, by XGBoost MinT-Shrink (53.7893%). The latter was not adopted as the selected model because doing so after viewing holdout performance would introduce selection bias. The weaker holdout result indicates temporal instability and reinforces the need for longer external evaluation, probabilistic intervals, and periodic model re-selection.

7. Visual Analysis of Total Sales Prediction

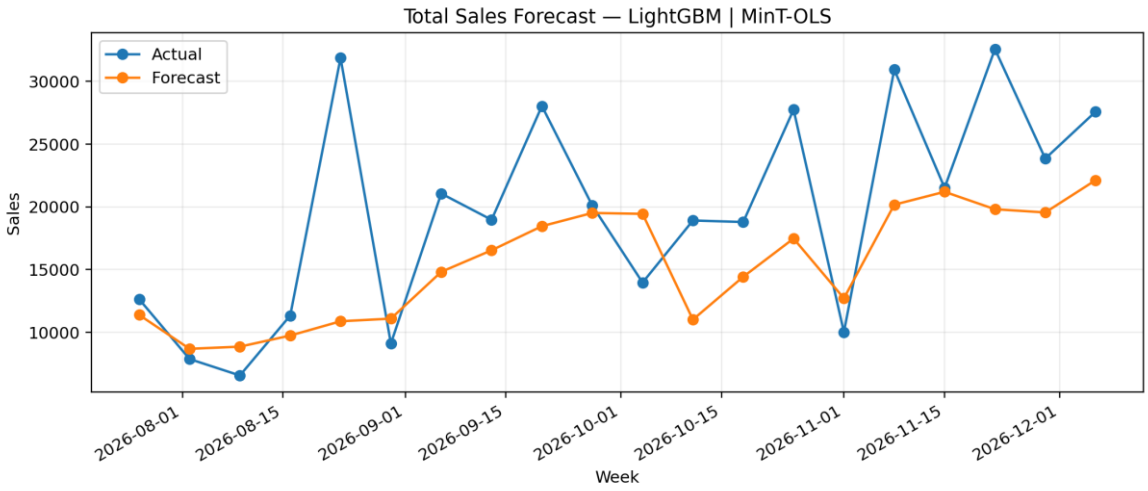


Figure 4. Comparison of actual and predicted total sales values using XGBoost MinT-OLS

Across the cross-validation windows, LightGBM MinT-OLS follows the medium-term level of total sales but smooths several sharp spikes and drops. The largest misses coincide with abrupt weekly changes that cannot be explained fully by lagged sales and calendar variables. Promotions, holidays, marketing campaigns, and price changes are absent from the dataset and are plausible omitted drivers.

8. Explainable AI Using SHAP

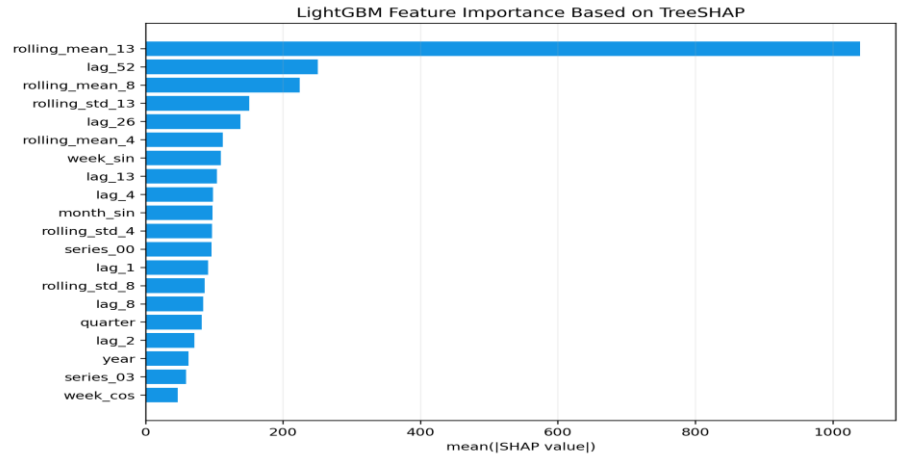


Figure 5. Feature importance of the selected LightGBM model based on mean absolute TreeSHAP values

TreeSHAP was calculated for the LightGBM base model refitted on all 205 development weeks; reconciliation is a linear post-processing step, so the attributions explain the base-model predictions rather than the MinT projection itself. The dominant feature was rolling_mean_13, followed by lag_52, rolling_mean_8, rolling_std_13, lag_26, rolling_mean_4, week_sin, lag_13, and lag_4. For managers, the rolling means provide demand baselines, lag_52 supports annual preparation, and rolling_std_13

Corresponding Authors : abdul.rohim.pranata18@gmail.com

identifies series requiring wider buffers or more frequent review. SHAP values describe model behavior, not causal effects.

Table 9. Interpretation of key features based on SHAP

Feature	Interpretation
rolling_mean_13	Trailing 13-week mean; medium-term demand baseline
lag_52	Sales in the corresponding week one year earlier
rolling_mean_8	Trailing eight-week demand level
rolling_std_13	Demand volatility over the previous 13 weeks
lag_26	Sales observed approximately six months earlier
rolling_mean_4	Short rolling mean over the previous four weeks
week_sin	Cyclical position within the annual weekly calendar
lag_13	Sales observed approximately one quarter earlier
lag_4	Sales observed four weeks earlier

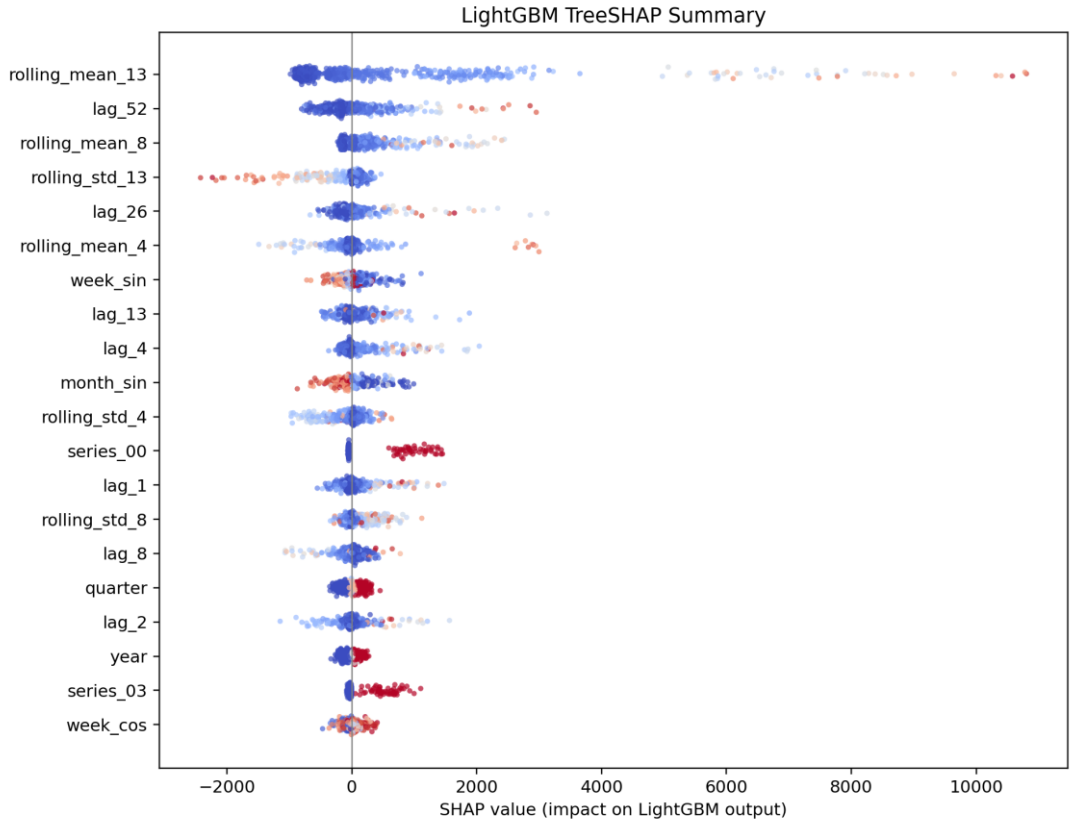


Figure 6. SHAP summary plot for the XGBoost model

The TreeSHAP summary shows that higher rolling_mean_13 and lag_52 values generally push predictions upward, whereas high rolling_std_13 values often contribute negatively. The series indicators separate aggregate from granular demand behavior, and calendar features provide smaller adjustments. Operationally, coherent forecasts allow the procurement total to equal category and subcategory allocations, reducing contradictory targets and double counting. Because the subcategory error remains high, coherent output should still be combined with service-level rules and exception-based inventory review.

Corresponding Authors : abdul.rohim.pranata18@gmail.com
Inkofar 2026 ©

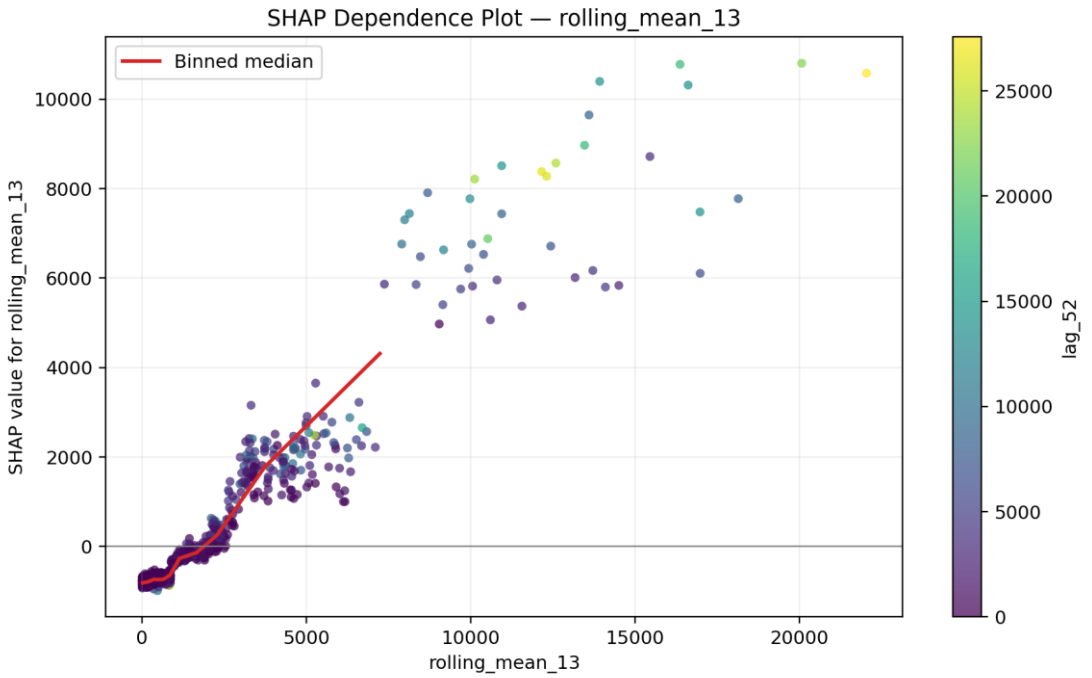


Figure 7. SHAP dependence plot for rolling_mean_13 in the selected LightGBM model

Figure 7 reveals a nonlinear relationship. Very low 13-week averages generally contribute negatively, values around 2,000-7,000 produce increasingly positive contributions, and high averages can add several thousand units to the prediction. The color scale indicates interaction with lag_52: observations with high annual-lag sales often occupy the upper positive region. The pattern supports using the 13-week mean as a planning baseline while consulting annual history before increasing replenishment commitments. The binned-median curve summarizes model response and should not be interpreted as a causal threshold

CONCLUSION

This study developed a reproducible explainable hierarchical retail-demand forecasting framework for one total, three category, and 17 subcategory weekly series. The corrected design reserves the last four weeks as a strict holdout and conducts five expanding-window rolling-origin folds exclusively on the preceding 205 weeks. LightGBM MinT-OLS was selected by cross-validation with a WAPE of 48.4886%, MAE of 1,344.3577, RMSE of 2,465.7518, MASE of 2.5970, and zero coherence error. Relative to LightGBM Base, reconciliation reduced cross-validation WAPE by 4.54% while eliminating aggregate contradictions. On the strict holdout, the selected configuration achieved a WAPE of 63.9589%; its weaker performance relative to Seasonal Naive during those four weeks cautions against claiming universal superiority.

The scientific contribution is the integration of leakage-safe model selection, explicit forecast reconciliation, per-series scale-aware evaluation, and TreeSHAP interpretation in one retail framework. Performance remained weakest for intermittent subcategories, so granular forecasts require safety-stock rules and managerial review. The SHAP dependence analysis showed that the 13-week rolling mean has a strong nonlinear influence moderated by annual history. Future work should evaluate longer external holdouts, promotions and holidays, Croston/SBA/TSB intermittent-demand benchmarks, probabilistic and cross-temporal reconciliation, and time-series foundation models under the same leakage-safe protocol. Validation on genuine operational data is required before deployment.

Corresponding Authors : *abdul.rohim.pranata18@gmail.com*

AUTHOR'S CONTRIBUTION

Abdul Rohim contributed to conceptualization, methodology design, model development, data analysis, and manuscript preparation. Novi Wulandari contributed to literature review, data interpretation, manuscript writing, and revision. Both contributed to validation, discussion development, manuscript review, and final approval of the article.

ACKNOWLEDGEMENTS

The authors express appreciation to their respective institutions for academic support during the preparation of this research manuscripty

DECLARATION OF COMPETING INTEREST

The authors declare that they have no competing interests.

FUNDING

This study did not receive any specific funding.

REFERENCES

1. Amiri, Y. I. A., & Wardati, N. K. (2025). Peramalan permintaan produk menggunakan ARIMA berbasis data mining. Jurnal Informatika: Jurnal Pengembangan IT, 10(3), 821–831. https://doi.org/10.30591/jpit.v10i3.8665
2. Arunika, K. R., & Dewi, L. J. E. (2025). Perbandingan model SARIMA, exponential smoothing, dan XGBoost untuk prediksi penjualan Super Store. Jurnal Informatika dan Teknik Elektro Terapan, 13(3S1). https://doi.org/10.23960/jitet.v13i3S1.8173
3. Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Panagiotelis, A. (2024). Forecast reconciliation: A review. International Journal of Forecasting, 40(2), 430–456. https://doi.org/10.1016/j.ijforecast.2023.10.010
4. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785
5. Gymnastiar, M., & Anwar, A. (2025). Penerapan model ARIMA dalam peramalan permintaan untuk meningkatkan efisiensi manajemen persediaan pada CV Kopi Biji Palembang. Jurnal Teknik Industri Terintegrasi, 8(4), 3926–3935. https://doi.org/10.31004/jutin.v8i4.48943
6. Hasanah, I. N., Lasarudin, A., & Hasyim, W. (2025). Penerapan data mining untuk prediksi penjualan produk terlaris menggunakan metode ANN dan LSTM. Jurnal Ilmu Komputer, 5(2). https://doi.org/10.31314/juik.v5i2.4180
7. Hidayat, K., Witanti, W., & Ramadhan, E. (2025). Analisis tren dan prediksi penjualan restoran menggunakan model time series Prophet. METIK Jurnal, 9(2), 457–467. https://doi.org/10.47002/metik.v9i2.1101
8. Jain, D. (2026). Superstore sales analysis [Data set]. Kaggle. https://www.kaggle.com/datasets/divyjain28/superstore-sales
9. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146-3154.
10. Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis, 88(2), 365-411. https://doi.org/10.1016/S0047-259X(03)00096-4

Corresponding Authors : abdul.rohim.pranata18@gmail.com

11. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.
12. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. PLOS ONE, 13(3), e0194889. https://doi.org/10.1371/journal.pone.0194889
13. Nailendra, S. Y., Witanti, W., & Abdillah, G. (2025). Optimasi prediksi penjualan retail online menggunakan LightGBM dan hyperparameter tuning. Jurnal Algoritma, 22(2), 1931–1942. https://doi.org/10.33364/algoritma/v.22-2.2551
14. Permana, M. H. (2024). Pengembangan model peramalan penjualan fast food dengan XGBoost menggunakan Python [Undergraduate thesis, Universitas Islam Negeri Syarif Hidayatullah Jakarta].
15. Putra, R. B. R., & Hendry. (2022). Multivariate time series forecasting pada penjualan barang retail dengan recurrent neural network. INOVTEK Polbeng: Seri Informatika, 7(1), 71–82.
16. Sunendar, N., & Rianto, Y. (2025). Comparison of ARIMA, LSTM, and GRU models for forecasting sales of HIT aerosol products. Pilar Nusa Mandiri, 21(2), 153–159. https://doi.org/10.33480/pilar.v21i2.6412
17. Suryawan, I. G. T., Putra, I. K. N., Meliana, P. M., & Sudipa, I. G. I. (2024). Performance comparison of ARIMA, LSTM, and Prophet methods in sales forecasting. Sinkron: Jurnal dan Penelitian Teknik Informatika, 8(4), 2410–2421. https://doi.org/10.33395/sinkron.v8i4.14057
18. Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. Journal of the American Statistical Association, 114(526), 804–819. https://doi.org/10.1080/01621459.2018.1448825

Corresponding Authors : abdul.rohim.pranata18@gmail.com