
Comparison of XGBoost and LightGBM for Product Demand Forecasting at KOPERASI XYZ

Muhammad Andhika Fahrezzy¹, Aries Dwi Indriyanti²

^{1,2} Universitas Negeri Surabaya, Surabaya, Indonesia

muhammadandhika.21045@mhs.unesa.ac.id, ariesdwiindriyanti@unesa.ac.id

ABSTRACT

Accurate product demand forecasting is essential for small internal retail units such as cooperative stores (koperasi), where manual, intuition-based stock decisions frequently cause overstock and stockouts. This study develops and compares two gradient boosting algorithms, XGBoost and LightGBM, to forecast daily product demand at KOPERASI XYZ, an internal store located in Sunter, North Jakarta. Historical transaction records from 1 January 2024 to 22 December 2025 were aggregated into daily demand series for four product categories: mineral water, tea drinks, cigarettes, and milk. A total of 37 predictive features were engineered, comprising lag features, rolling and expanding statistics, and calendar features that encode 34 Indonesian national holidays. The data were divided using a chronological (time-based) split of 60:20:20, and both models were tuned using Grid Search cross-validation with a log-transformed target. Forecasting performance was evaluated using the Root Mean Squared Error (RMSE) and the Mean Absolute Percentage Error (MAPE). The results show that LightGBM outperforms XGBoost, achieving an overall RMSE of 24.53 and a MAPE of 9.72% (highly accurate), compared with an RMSE of 27.52 and a MAPE of 11.75% for XGBoost, corresponding to improvements of 10.86% in RMSE and 17.28% in MAPE. The best model was integrated into an interactive Streamlit dashboard that visualizes forecasts and recommends monthly stock levels to support data-driven inventory planning.

Keywords: Demand Forecasting, XGBoost; LightGBM, Gradient Boosting, Time Series, Inventory Management

Article Info:

Article history:

Received 04 07, 2026

Revised 05 13, 2026

Accepted 06 11, 2026

Corresponding Author

Muhammad Andhika Fahrezzy

Universitas Negeri Surabaya, Surabaya, Indonesia

muhammadandhika.21045@mhs.unesa.ac.id

1. INTRODUCTION

Demand forecasting is a fundamental component of modern retail and supply-chain management. Reliable estimates of future demand allow businesses to plan procurement, optimize inventory levels, reduce operating costs, and improve customer satisfaction [1], [2]. When demand is underestimated, stores experience stockouts and lost sales; when it is overestimated, capital is tied up in excess inventory and, for perishable goods, products may expire before they are sold. Balancing these two risks is therefore central to profitable retail operations.

KOPERASI XYZ is an internal cooperative retail store located in Sunter, North Jakarta, that serves the daily needs of its members and employees. Like many small retail units, its stock replenishment decisions have traditionally been made manually, relying on the intuition and experience of staff rather than on quantitative analysis of historical sales. This approach becomes increasingly unreliable as the number of products grows and as demand fluctuates due to

weekdays, weekends, paydays, and national holidays. As a result, the store frequently faces both overstock of slow-moving items and stockouts of fast-moving ones.

The products sold at KOPERASI XYZ are dominated by fast-moving consumer goods such as mineral water, tea drinks, cigarettes, and milk. These categories exhibit different demand patterns and volatilities: some sell in large, highly variable volumes, while others sell in small but steady quantities. Capturing these heterogeneous patterns, together with strong calendar and holiday effects specific to Indonesia, requires a forecasting method that can learn complex non-linear relationships from historical data rather than relying on fixed statistical assumptions.

Classical time-series methods such as ARIMA and exponential smoothing have been widely used for demand forecasting, but they often struggle with multiple interacting seasonalities, irregular holiday effects, and non-linear dynamics [3]. Machine-learning approaches have consequently gained popularity because they can incorporate rich engineered features and model non-linear behaviour [2], [4]. Among them, gradient boosting decision-tree algorithms, in particular XGBoost [5] and LightGBM [6], have repeatedly delivered state-of-the-art accuracy in large-scale forecasting benchmarks such as the M4 and M5 competitions [9], [10], while remaining computationally efficient and interpretable through feature-importance analysis [8], [12].

Within this line of work, various techniques have been applied to sales data of small businesses in Indonesia, ranging from simple statistical baselines such as the Simple Moving Average, which was used to forecast cookie sales at a small food enterprise [18], to interpretable decision-tree classifiers such as C4.5, which was used to distinguish best-selling from slow-moving promotional products [19]. However, these studies rely either on moving-average smoothing or on categorical classification, rather than on modern gradient-boosting regressors that can jointly model lag, rolling, and calendar effects.

Despite this progress, the application of gradient boosting to small internal cooperative retail stores remains limited. Most published work focuses on large retailers or open benchmark datasets, and relatively few studies combine daily category-level forecasting with an extensive set of engineered calendar features that reflect the Indonesian national-holiday calendar. Furthermore, few studies translate the resulting model into a practical decision-support tool that non-technical staff can operate. This study addresses that gap by comparing XGBoost and LightGBM for daily product demand forecasting at KOPERASI XYZ and by deploying the best model in an interactive dashboard.

The objectives of this research are threefold: (1) to build daily demand-forecasting models for four product categories at KOPERASI XYZ using XGBoost and LightGBM; (2) to compare the two algorithms using the RMSE and MAPE metrics in order to identify the more accurate model; and (3) to integrate the best-performing model into an interactive dashboard that recommends monthly stock levels. The main contributions are a comprehensive feature-engineering scheme comprising 37 lag, rolling, and calendar features that encode 34 Indonesian national holidays; an objective, time-aware comparison of two leading gradient-boosting algorithms on real cooperative-store data; and a deployable Streamlit dashboard that turns the forecasts into actionable stock recommendations.

2. RESEARCH METHOD

This study applies a quantitative, experimental approach in which two supervised machine-learning models are trained and evaluated on historical sales data. The overall workflow, from problem identification to dashboard implementation, is organized into five sequential phases as summarized in Figure 1.

2.1 Research Stages

The research was carried out in five phases. The preparation phase covers problem identification, definition of the research objectives, and collection of historical sales data from KOPERASI XYZ. The data-processing phase includes preprocessing, exploratory data analysis (EDA), feature engineering, and chronological dataset splitting. The modeling phase covers model building with XGBoost and LightGBM and hyperparameter optimization using Grid Search cross-validation. The evaluation and analysis phase measures model accuracy with RMSE and MAPE and compares the two algorithms to select the best model. Finally, the dashboard-implementation phase integrates the selected model into an interactive application that visualizes forecasts and supports stock planning. These stages are shown in Figure 1.

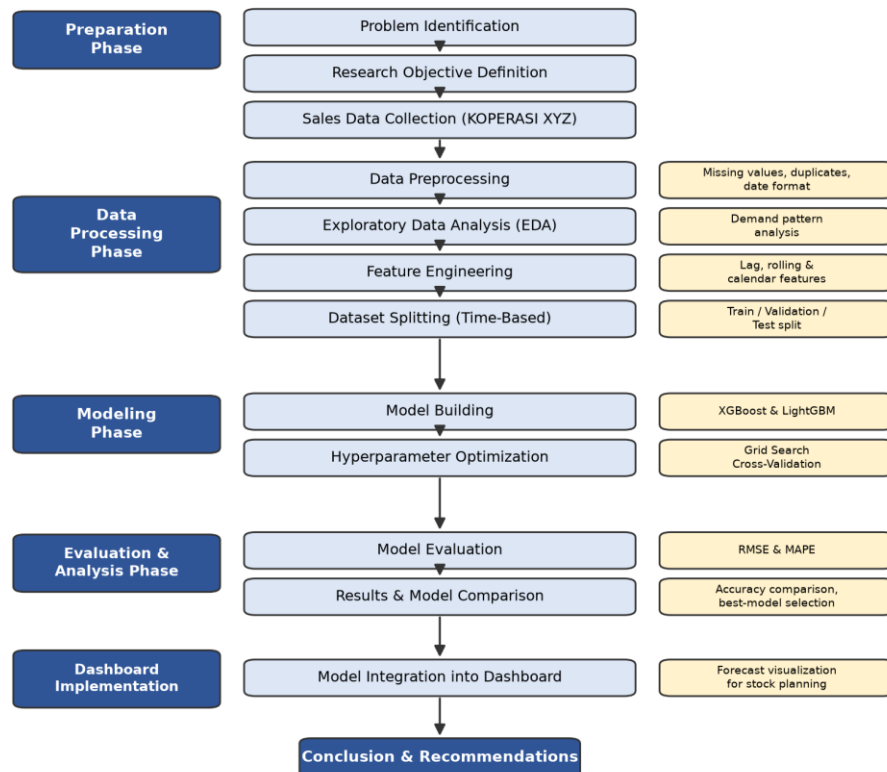


Figure 1. Research stages of the study.

2.2 Data Collection and Description

The dataset consists of historical transaction records from KOPERASI XYZ covering the period from 1 January 2024 to 22 December 2025. The raw data contain approximately 45,200 transaction rows across many products. For this study, the analysis focuses on four representative fast-moving categories: mineral water, tea drinks, cigarettes, and milk. Figure 2 illustrates the aggregated monthly sales trend of these categories, which shows clear fluctuations across the observation period and motivates the use of a flexible, feature-based forecasting model.

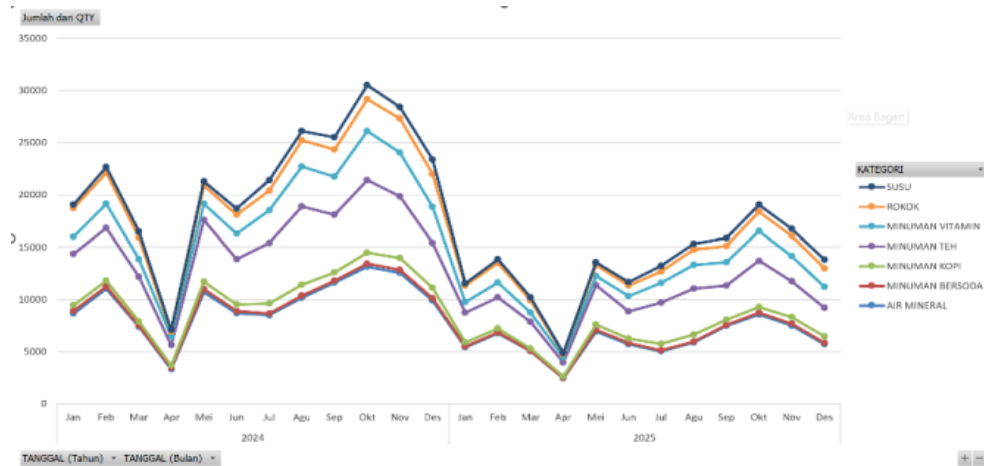


Figure 2. Monthly sales trend of the four product categories.

2.3 Data Preprocessing

Preprocessing prepared the raw transactions for modeling. First, transactions that did not belong to the four target categories were removed, eliminating 14,073 rows and leaving 31,107 relevant records. Duplicate rows were dropped, inconsistent date formats were standardized, and missing values were checked and handled. The cleaned transactions were then aggregated to a daily granularity per category, producing a continuous daily demand series of 2,880 category-day observations. This chronologically ordered daily series forms the basis for feature engineering and modeling.

2.4 Feature Engineering

Because tree-based models cannot infer temporal order on their own, 37 predictive features were engineered to expose the temporal, statistical, and calendar structure of the series to the models. Lag features capture the value of demand at previous time steps, as defined in (1), where $y(t)$ is the demand at day t and k is the lag length.

$$LAG_k(t) = y(t - k), \quad k \in \{1, 2, 3, 7, 14, 28\} \quad (1)$$

Rolling and expanding statistics summarize recent demand behaviour. A rolling mean over a window of length p is defined in (2); rolling standard deviations and expanding statistics are computed analogously to describe short- and long-term variability.

$$RM_p(t) = (1/p) \sum_{j=1..p} y(t - j) \quad (2)$$

Calendar features encode the day, day of week, week, month, quarter, year, day of year, and a weekend indicator, together with binary indicators for 34 Indonesian national holidays observed during 2024 and 2025. The complete feature set is summarized in Table 1.

Table 1. Summary Of Engineered Features

Feature Group	Count	Description / Examples
Lag features	6	Demand at t-1, t-2, t-3, t-7, t-14, t-28
Rolling & expanding statistics	8	Rolling mean & std (windows 3, 7, 14) and expanding mean & std
Calendar features	23	Day, day-of-week, week, month, quarter, year, day-of-year, weekend flag, and 34 Indonesian holiday indicators
Total	37	Combined predictive feature set

2.5 Dataset Splitting

To respect the temporal nature of the data and avoid look-ahead bias, a chronological (time-based) split of 60:20:20 was used instead of random shuffling. The earliest observations were used for training, the following period for validation and hyperparameter tuning, and the most recent

period for testing. Figure 3 illustrates this time-based partitioning, and Table 2 details the exact periods and sizes of each subset.



Figure 3. Illustration of the time-based (chronological) data split.

Table 2. Time-Based Data Split

Subset	Period	Days	Rows
Training	10 Feb 2024 - 12 May 2025	269	1,076
Validation	15 May 2025 - 24 Aug 2025	68	272
Testing	25 Aug 2025 - 22 Dec 2025	85	340
Total	10 Feb 2024 - 22 Dec 2025	422	1,688

2.6 Forecasting Algorithms

Two gradient boosting decision-tree algorithms were compared. Gradient boosting builds an ensemble of decision trees sequentially, where each new tree is trained to correct the residual errors of the previous ensemble, as illustrated in Figure 4 [8]. XGBoost optimizes a regularized objective function, shown in (3), where l is the loss between the actual value y_i and the prediction, and the regularization term penalizes model complexity to reduce overfitting [5].

$$Obj = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k), \quad \Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (3)$$

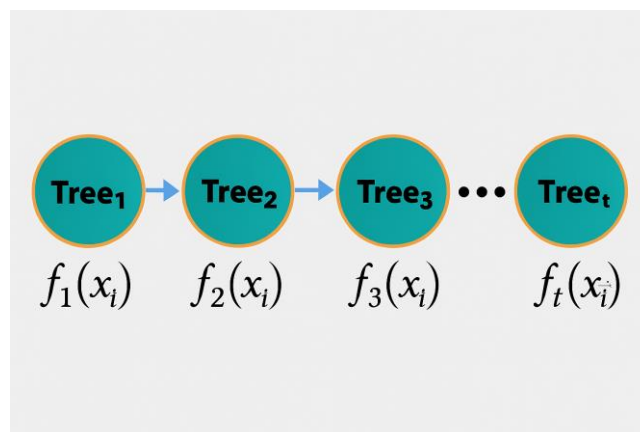


Figure 4. Sequential tree building in gradient boosting.

XGBoost grows trees level-wise (depth-wise), whereas LightGBM grows trees leaf-wise, splitting the leaf with the largest loss reduction. LightGBM also uses histogram-based binning of features, which makes it faster and often more accurate on large feature sets [6]. The difference between the two tree-growth strategies is illustrated in Figure 5.

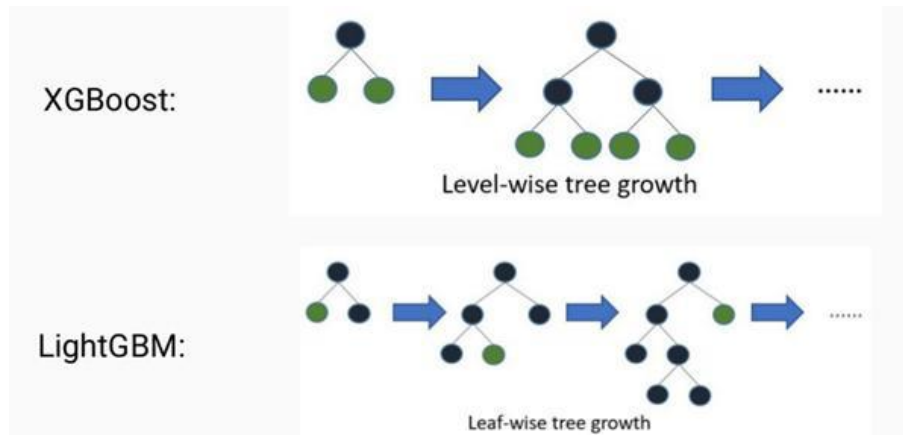


Figure 5. Level-wise (XGBoost) versus leaf-wise (LightGBM) tree growth.

2.7 Hyperparameter Optimization and Target Transformation

Both models were tuned using Grid Search with cross-validation on the training and validation data to select the combination of hyperparameters that minimizes the validation error. Because daily demand is right-skewed, the target variable was transformed using the log1p transformation defined in (4) and wrapped in a TransformedTargetRegressor, so that predictions are automatically returned to the original scale. This stabilizes variance and improves accuracy for categories with occasional large spikes.

$$y' = \log(1 + y), \quad \hat{y} = \exp(\hat{y}') - 1 \quad (4)$$

The best hyperparameter configurations obtained for each model are listed in Table 3.

Table 3. Best Hyperparameter Configuration

Hyperparameter	XGBoost	LightGBM
n_estimators	100	200
max_depth / num_leaves	2	7
learning_rate	0.05	0.05
subsample	0.9	0.9
colsample_bytree	0.9	0.9
min_child_weight / min_child_samples	1.0	5

2.8 Evaluation Metrics

Model accuracy was measured using two complementary metrics. The Root Mean Squared Error (RMSE), (5), expresses the average magnitude of the errors in the original units and penalizes large errors more heavily. The Mean Absolute Percentage Error (MAPE), (6) expresses the error as a percentage and allows comparison across categories with different demand scales [7].

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad (5)$$

$$MAPE = \frac{100\%}{n} \sum |(y_i - \hat{y}_i) / y_i| \quad (6)$$

To interpret MAPE values, the widely used scale proposed by Lewis [11] was adopted, as shown in Table 4.

Table 4. Mape Interpretation Scale

MAPE Range	Interpretation
< 10%	Highly accurate forecasting
10% - 20%	Good forecasting
20% - 50%	Reasonable forecasting
> 50%	Inaccurate forecasting

2.9 Implementation Environment

The experiments were implemented in Python using the xgboost (v3.2.0) and lightgbm (v4.6.0) libraries, together with pandas and numpy for data manipulation and joblib for model persistence [13], [14]. The final model was deployed in an interactive dashboard built with Streamlit and Altair for visualization.

3. RESULTS AND DISCUSSION

3.1 Dataset Overview and Exploratory Analysis

Exploratory analysis of the daily demand series revealed clear differences between the four categories. Mineral water is a high-volume, highly variable category, with an average of roughly 410 units sold per day, a maximum above 1,200 units, and a standard deviation of about 187.51 units. Milk, in contrast, is a low-volume category with an average of only about 33.87 units per day. Tea drinks and cigarettes fall between these extremes. Figure 6 shows the daily demand series of the four categories, highlighting strong day-to-day fluctuations as well as recurring weekly and holiday-related patterns.

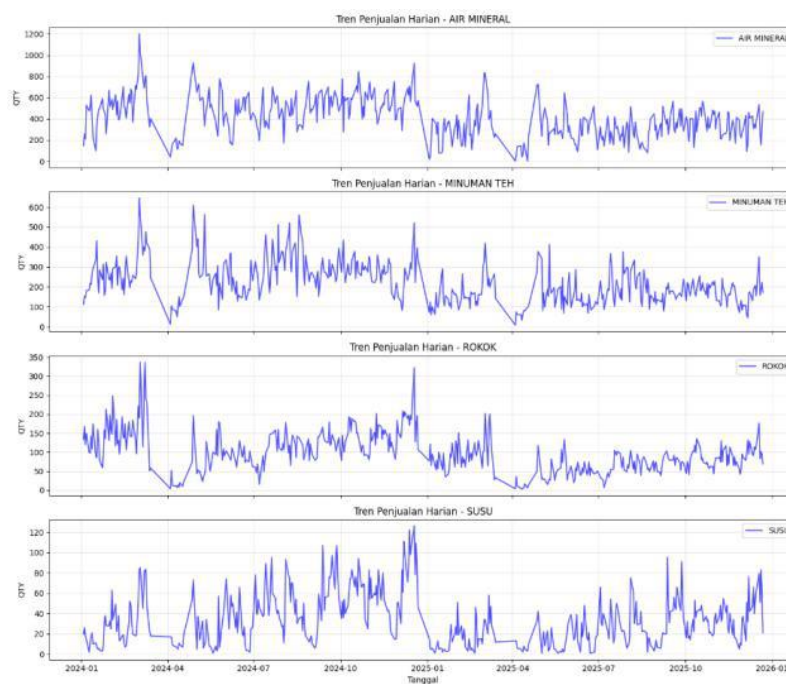


Figure 6. Daily demand series for the four product categories.

After feature engineering, the observations that contained undefined lag or rolling values (the initial warm-up period) were removed, resulting in a final modeling dataset of 1,688 rows corresponding to 422 effective days across the four categories. This dataset was used for all training, validation, and testing.

3.2 Hyperparameter Optimization Results

Grid Search cross-validation selected relatively shallow XGBoost trees (max_depth of 2) with 100 estimators, and slightly more expressive LightGBM trees (num_leaves of 7) with 200 estimators, both using a small learning rate of 0.05 and 90% subsampling of rows and columns, as previously listed in Table III. The shallow trees and strong subsampling act as regularization, which is appropriate given the limited number of training days and helps prevent overfitting to short-term noise.

3.3 Overall Model Comparison

On the held-out test set, LightGBM consistently outperformed XGBoost. Across all categories, LightGBM achieved an overall RMSE of 24.53 and a MAPE of 9.72%, which falls in the highly accurate range according to Table 5. XGBoost achieved an overall RMSE of 27.52 and a MAPE of 11.75%, which is classified as good. In relative terms, LightGBM reduced the RMSE by 10.86% and the MAPE by 17.28% compared with XGBoost. The per-category results are presented in Table V and visualized in Figure 7 and Figure 8.

Table 5. Rmse And Mape Comparison Per Category

Category	RMSE (XGB)	RMSE (LGB)	MAPE (XGB)	MAPE (LGB)
Mineral Water	50.71	46.73	11.87%	11.24%
Tea Drinks	18.86	13.49	8.81%	7.83%
Cigarettes	9.20	4.98	11.13%	4.99%
Milk	4.12	3.97	15.20%	14.81%
All Categories	27.52	24.53	11.75%	9.72%

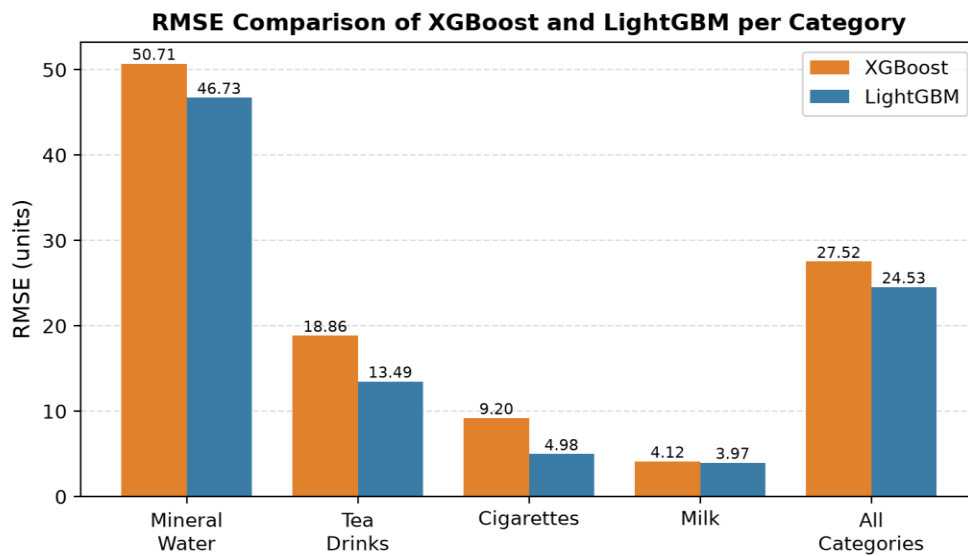


Figure 7. RMSE comparison of XGBoost and LightGBM per category.

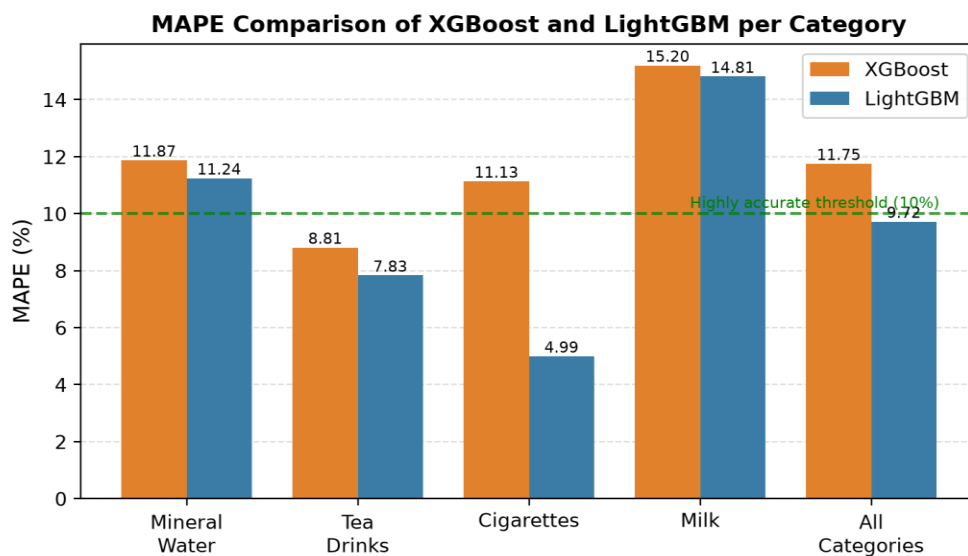


Figure 8. MAPE comparison of XGBoost and LightGBM per category.

LightGBM achieved a lower RMSE and MAPE than XGBoost in every category, confirming that its leaf-wise growth and histogram-based splitting are better suited to this dataset. The largest relative improvement occurs for cigarettes, where the MAPE drops from 11.13% with XGBoost to only 4.99% with LightGBM.

3.4 Per-Category Performance of LightGBM

Table 6 summarizes the accuracy of the selected LightGBM model per category together with the demand characteristics that explain the results. Cigarettes and tea drinks are forecast with very high accuracy (MAPE below 8%) because their demand is relatively stable and predictable. Mineral water, despite its large absolute RMSE, is still forecast well in percentage terms (MAPE of 11.24%) given its high and volatile volume. Milk shows the highest MAPE (14.81%) because its very low daily volume means that even small absolute errors translate into large percentage errors, yet its absolute RMSE remains the smallest of all categories.

Table 6. Lightgbm Performance By Category

Category	MAPE	Volume Characteristic	Interpretation
Cigarettes	4.99%	Medium volume, stable	Highly accurate
Tea Drinks	7.83%	Medium volume, predictable	Highly accurate
Mineral Water	11.24%	High volume, large fluctuation	Good
Milk	14.81%	Low volume, small absolute error	Good

3.5 Feature Importance Analysis

Feature-importance analysis provides insight into which predictors drive the forecasts. For XGBoost, the most influential features are the short-term rolling means, with ROLLING_MEAN_3 contributing about 29.71% and ROLLING_MEAN_7 about 23.38% of the total importance, indicating that recent average demand is the strongest signal. For LightGBM, the first lag (LAG_1) dominates the top splits, followed by calendar features such as the day of the month and the day of the year. In both models, the combination of recent-history and calendar features confirms that the engineered features successfully captured the temporal and seasonal structure of demand.

3.6 Interactive Forecasting Dashboard

To make the results usable by store staff, the selected LightGBM model was integrated into an interactive dashboard built with Streamlit. The dashboard provides an overview page that summarizes total historical sales and the latest forecast accuracy, as shown in Figure 9, and a forecasting page that generates future demand and recommended stock levels, as shown in Figure 10.

Using the deployed model, the dashboard reported a total of 342,850 units of actual historical sales and produced a recommended total stock of 7,561 units for January 2026, of which mineral water accounts for the largest share at about 67.60%. The dashboard also allows results to be exported to Excel, so that staff can integrate the recommendations directly into their purchasing process.

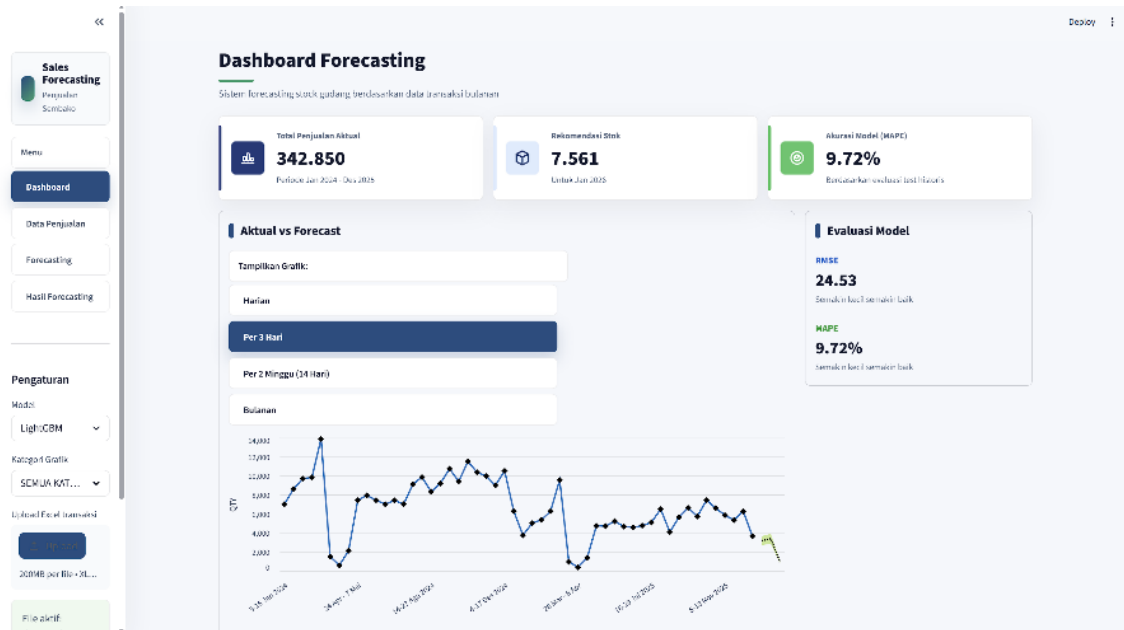


Figure 9. Dashboard overview page.

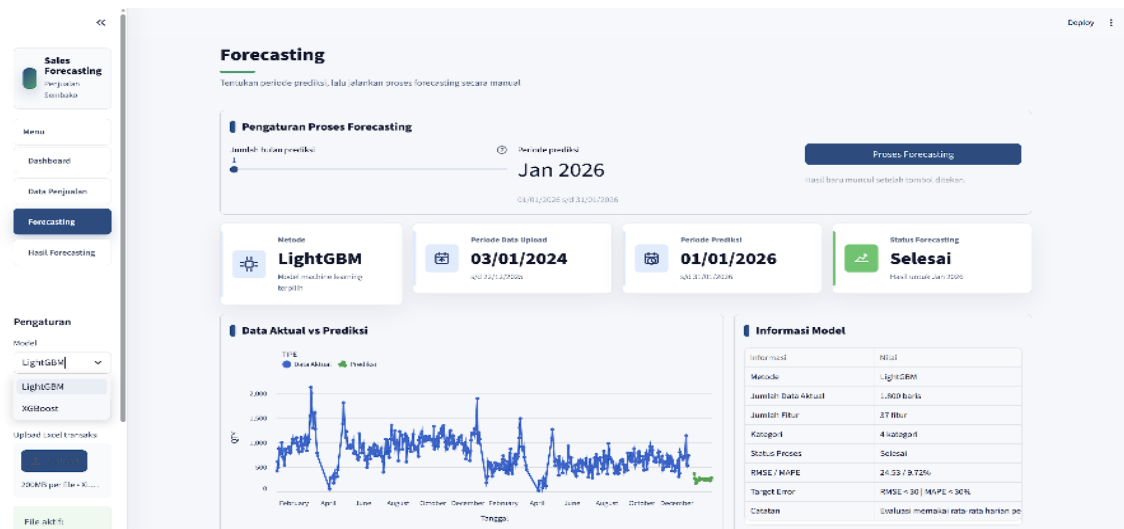


Figure 10. Forecasting page with recommended stock levels.

3.7 Discussion and Managerial Implications

The results demonstrate that gradient boosting, and LightGBM in particular, can accurately forecast daily demand for a small internal cooperative store when combined with careful feature engineering and a time-aware evaluation. From a managerial perspective, replacing intuition-based ordering with model-driven stock recommendations can reduce both overstock and stockouts, free up working capital, and lower the risk of spoilage for perishable items. Because the model runs behind a simple dashboard, staff without a technical background can obtain reliable forecasts and act on them directly.

3.8 Comparison with Prior Studies

The findings are consistent with prior research reporting the strong performance of gradient boosting for retail and sales forecasting [4], [9], [10]. The observed advantage of LightGBM over XGBoost on a tabular, feature-rich dataset also agrees with results from large-scale benchmarks. This study extends that body of work by focusing on a small internal cooperative store, by

encoding the Indonesian national-holiday calendar, and by delivering an end-to-end decision-support tool rather than an offline model only.

3.9 Limitations and Future Work

This study has several limitations. The analysis is restricted to four product categories from a single store, so the results may not generalize directly to other products or locations. The observation period, while covering two years, may not capture longer-term structural changes in demand. Future work could extend the approach to a larger product catalogue, incorporate external variables such as promotions, weather, and price, and explore hybrid or deep-learning models. Automated model retraining and monitoring within the dashboard would also help maintain accuracy as demand patterns evolve.

CONCLUSION

This study developed and compared XGBoost and LightGBM for daily product demand forecasting at KOPERASI XYZ, an internal cooperative retail store, using two years of historical sales data for four fast-moving categories. A comprehensive set of 37 lag, rolling, and calendar features, including 34 Indonesian national-holiday indicators, was engineered, and both models were tuned with Grid Search cross-validation using a time-based split. The results show that LightGBM outperforms XGBoost across all categories, achieving an overall RMSE of 24.53 and a MAPE of 9.72% (highly accurate) versus 27.52 and 11.75% for XGBoost, an improvement of 10.86% in RMSE and 17.28% in MAPE. The best model was deployed in an interactive Streamlit dashboard that visualizes forecasts and recommends monthly stock levels, converting the analysis into a practical decision-support tool. These results confirm that LightGBM is an accurate and efficient choice for daily demand forecasting in small retail settings and that data-driven forecasting can meaningfully improve inventory planning at KOPERASI XYZ.

ACKNOWLEDGEMENTS

The authors thank KOPERASI XYZ for providing the sales data used in this study and the Information Systems study program, Universitas Negeri Surabaya, for its academic support.

REFERENCES

- [1] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Melbourne, Australia: OTexts, 2021.
- [2] R. Carboneau, 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.
- [3] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159-175, 2003.
- [4] B. M. Pavlyshenko, "Machine-learning models for sales time series forecasting," *Data*, vol. 4, no. 1, p. 15, 2019.
- [5] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785-794.
- [6] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017, pp. 3146-3154.

- [7] 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.
- [8] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189-1232, 2001.
- [9] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M4 competition: 100,000 time series and 61 forecasting methods," *International Journal of Forecasting*, vol. 36, no. 1, pp. 54-74, 2020.
- [10] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M5 accuracy competition: Results, findings and conclusions," *International Journal of Forecasting*, vol. 38, no. 4, pp. 1346-1364, 2022.
- [11] C. D. Lewis, *Industrial and Business Forecasting Methods*. London, U.K.: Butterworth Scientific, 1982.
- [12] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers in Neurorobotics*, vol. 7, p. 21, 2013.
- [13] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.
- [14] W. McKinney, "Data structures for statistical computing in Python," in *Proc. 9th Python in Science Conf. (SciPy)*, 2010, pp. 56-61.
- [15] N. K. Ahmed, A. F. Atiya, N. E. Gayar, and H. El-Shishiny, "An empirical comparison of machine learning models for time series forecasting," *Econometric Reviews*, vol. 29, no. 5-6, pp. 594-621, 2010.
- [16] J. Brownlee, *Deep Learning for Time Series Forecasting. Machine Learning Mastery*, 2018.
- [17] Streamlit Inc., "Streamlit: A faster way to build and share data apps," 2024. [Online]. Available: <https://streamlit.io>
- [18] K. Widi and A. D. Indriyanti, "Peramalan penjualan cookies pada usaha cookies Sweetnest menggunakan metode Simple Moving Average," *Journal of Emerging Information Systems and Business Intelligence*, vol. 5, no. 2, pp. 104-109, 2024.
- [19] R. S. Iswandar and A. D. Indriyanti, "Sales performance classification of promotional products using data mining," *Journal of Emerging Information Systems and Business Intelligence*, vol. 7, no. 2, pp. 213-220, 2026.