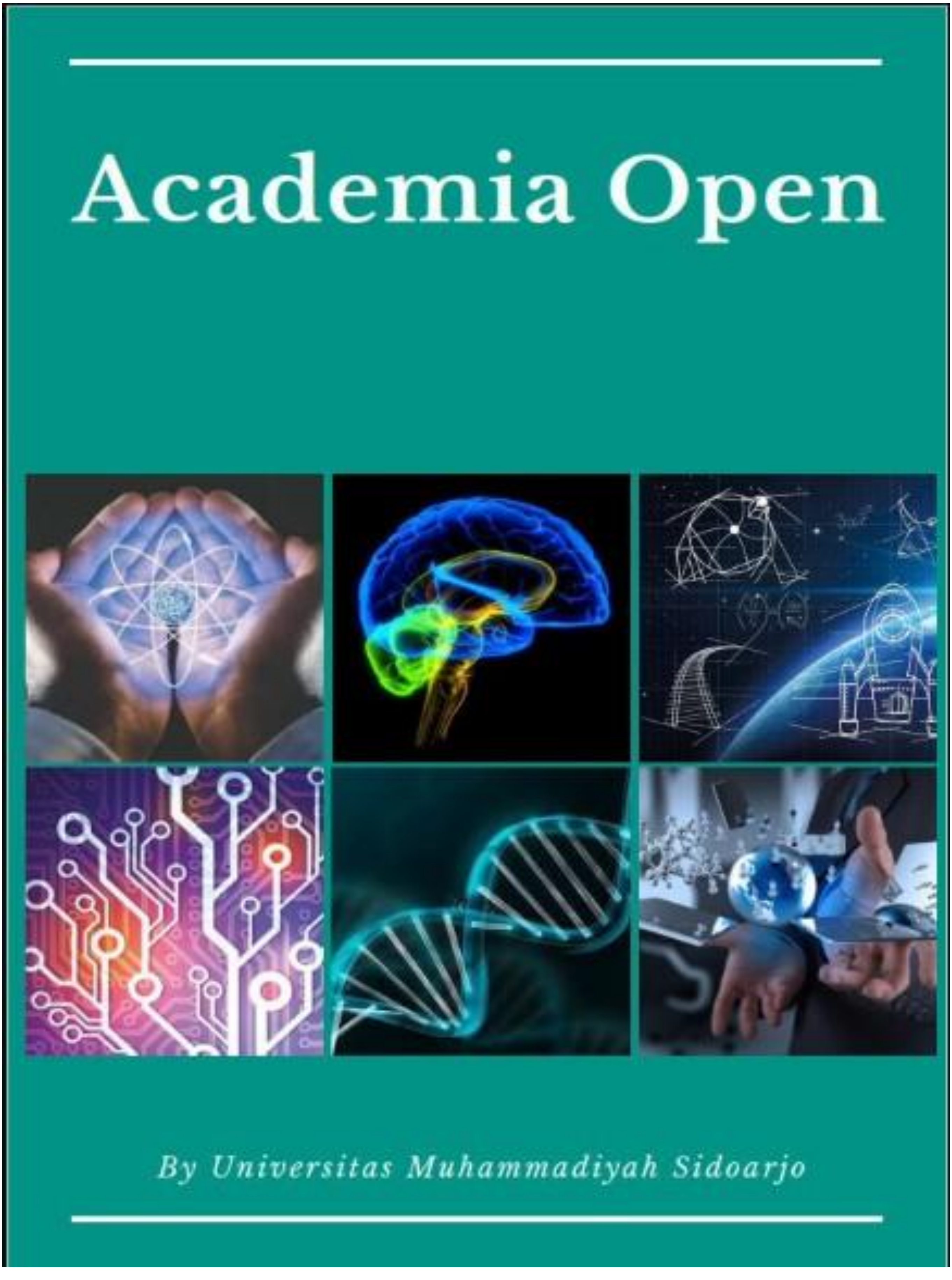
Academia Open

Vol. 11 No. 1 (2026): June
DOI: 10.21070/acopen.11.2026.14690



ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Academia Open

Vol. 11 No. 1 (2026): June
DOI: 10.21070/acopen.11.2026.14690

Table Of Contents

Academia Open
Vol. 11 No. 1 (2026): June
DOI: 10.21070/acopen.11.2026.14690

Originality Statement

The author[s] declare that this article is their own work and to the best of their knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the published of any other published materials, except where due acknowledgement is made in the article. Any contribution made to the research by others, with whom author[s] have work, is explicitly acknowledged in the article.

Conflict of Interest Statement

The author[s] declare that this article was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright Statement

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Academia Open

Vol. 11 No. 1 (2026): June
DOI: 10.21070/acopen.11.2026.14690

EDITORIAL TEAM

Academia Open
Vol. 11 No. 1 (2026): June
DOI: 10.21070/acopen.11.2026.14690

Article information

Check this article update (crossmark)

Check this article impact (*)

Save this article to Mendeley

(*) Time for indexing process is various, depends on indexing database platform

Academia Open

Vol. 11 No. 1 (2026): June
DOI: 10.21070/acopen.11.2026.14690

Stacking Ensemble Yields Lower Drinking Gallon Demand Forecasting Error

Faldo Julian Joshua Laoh, faldolaoh357@gmail.com (*)

Program Studi Sistem Informasi, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia

Prasasti Karunia Farista Ananto, prasasti.karunia.fasilkom@upnjatim.ac.id

Program Studi Sistem Informasi, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia

Mohammad Al Hafidz, hafidz.si@upnjatim.ac.id

Program Studi Sistem Informasi, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia

(*) Corresponding author

Abstract

General Background Refill drinking water depots areessential for Indonesian communities, yetthey often suffer from operational inefficiencies due to reactive inventory management. **Specific Background** Depot XYZ experiences highly volatile daily demand, frequently resulting in stockouts and service delays. **Knowledge Gap** The optimal forecasting approach for daily gallon demand in this sector remains underexplored. **Aims** This study develops a demand forecasting model utilizing SARIMA and Prophet as base learners, optimized through an XGBoost-based stacking ensemble. **Results** The stacking ensemble achieved the lowest forecasting error, with an RMSE of 73.648907 and a MAPE of 18.32%, outperforming standalone models. **Novelty** Results demonstrate that ensemble accuracy is contingent upon the complementary information provided by base learners, noting marginal improvements when a dominant statistical model already captures core temporal patterns. **Implications** This framework enables proactive inventory planning, helping depot management mitigate stockout risks and ensure consistent service fulfillment.

Highlights:

- Integration of SARIMA and Prophet base learners within an XGBoost meta-learning framework minimizes daily replenishment errors.
- Empirical evaluation confirms that combined machine learning architectures provide superior predictive precision compared to isolated statistical models.
- Practical utility of ensemble forecasting enables proactive inventory management to mitigate operational disruptions and service fulfillment delays.

Keywords: Demand Forecasting, SARIMA, Prophet, Stacking Ensemble, XGBoost

Published date: 2026-06-26

Academia Open
Vol. 11 No. 1 (2026): June
DOI: 10.21070/acopen.11.2026.14690

I. Introduction

The refill drinking water depot industry (DAMIU) has become an essential service in Indonesia, supported by high community dependence on refill drinking water. Based on data from the ministry of industry in collaboration with the Indonesian Refill Water Depot Association (ASDAMINDO), there are more than 78,000 depot units operating nationally, while the Central Statistics Agency (BPS) reports that around 34,49% of households rely on refill water as their main drinking water source [1], [2]. This indicates that depot operations must ensure continuous availability and reliable fulfillment of daily demand.

As part of this ecosystem, Depot XYZ in Balikpapan operates a hybrid B2C and B2B business model with an average daily demand of around 280 gallons. However, demand patterns are highly fluctuating and do not show fully stable seasonal behavior. In operational practice, inventory management is still reactive, where stock replenishment is only performed after depletion occurs during ongoing orders, leading to delays in service fulfillment.

This reactive system has resulted in operational inefficiencies, including 40 recorded out-of-stock incidents over a three-year period. Delivery delays can reach up to several hours depending on production condition while normal lead time is only around 10-50 minutes. With the average yearly out-of-stock (OOS) frequency is 6,67%, this condition is considered significant when OOS reaching more than 4,94% in operational studies [3]. These disruptions contribute to service instability and customer dissatisfaction.

To address this issue, demand forecasting is required as data-driven approach for predicting future demand based on historical patterns. Time series forecasting enables proactive inventory planning so that stock replenishment can be scheduled before demand spikes occur. In this study, SARIMA and Prophet are selected as base models due to their capability in capturing autocorrelation, seasonality, trend, and irregular patterns in demand data [4], [5].

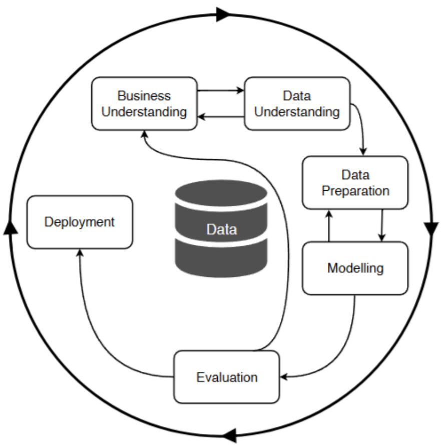
However, No single forecasting model is consistently optimal across all data characteristics. Recent studies show that ensemble approaches can empirically improve prediction stability by combining complementary information from multiple models [6]. Stacking ensemble has also been applied in operational forecasting contexts, such as food-demand forecasting and primary energy consumption, where prediction improvement obtained by learning complementary patterns from different base models[6], [7]. However, its use for daily drinking gallon demand forecasting in refill drinking water depot operations remains underexplored. Therefore, this study proposes a stacking ensemble using SARIMA and Prophet as base learners and XGBoost as a meta-learner, with the aim of improving forecasting accuracy and supporting inventory threshold decisions. The study focuses on daily demand data from October 2023 to Januari 2026 and evaluates performance using RMSE and MAPE [8], [9].

II. Method

A. Research Framework

This study adopts the Cross-Industry Standard Process for Data Mining, also known as CRISP-DM as the main methodological approach. CRISP-DM provides an iterative and structured workflow consisting of business understanding, data understanding, data preparation, modeling, evaluation, and deployment as pictured on the Figure 1 below [10]. Since this article focuses specifically on the forecasting process, the discussion is limited to the first five phases. The deployment phase of any form such as website or inventory implementation is outside the scope of this article.

Figure 1. The flow stages of the CRISP-DM framework [10]



B. Business Understanding

The business understanding phase identifies the operational issue faced by Depot XYZ. The depot still applies a reactive replenishment mechanism, in which stock preparation is conducted after inventory shortages occur during daily operations. This condition may increase the risk of stockout and delay order fulfillment, particularly during sudden increase in demand. Stockout prediction and inventory planning are relevant operational issues because insufficient stock availability may disrupt

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

service performance and reduce the ability of a business to fulfill customer demand [11].

Based on this problem, the analytical objective of this study is to develop a daily gallon demand forecasting model, SARIMA dan Prophet are used as base learners, while XGBoost is applied as a meta-learner to optimize the prediction outputs generated by both models.

C. Data Understanding

The dataset consists of historical daily gallon demand records from Depot XYZ in Indonesia. The observation period covers October 1st, 2023, to January 22nd, 2026. The final dataset contains 845 daily observations, which are sufficient to represent historical demand variation across more than two years of depot operations. The dataset is structured as a univariate time series with two primary variables: the observation date and the total daily demand.

Tie series analysis requires initial understanding of trend, seasonality, and irregular variation because these components influence model selection and forecasting accuracy [9]. Therefore, exploratory analysis is conducted to identify temporal patterns, daily fluctuations, missing dates, and potential weekly seasonality before the modeling process. Thus, the main variables used in this research is as shown on the Table 1 below.

Table 1. Research Variables

Variable	Description
tanggal_transaksi	Date of daily observation
qty	Daily gallon demand (pcs)

D. Data Preparation

The data preparation phase transforms raw operational records into a consistent daily time series format. First, the transaction date is converted into a standardized datetime format. Second, the observations are sorted chronologically top reserve their temporal order. Third, a continuous daily index is constructed to identify dates without recorded demand values. Missing values generated during this process are handled using temporal interpolation to maintain data continuity.

The processed dataset is divided chronologically into training and testing sets using an 80:20 ratio. The first 80% of the observations are used for model training, while the remaining 20% are used for final out-of-sample evaluation. A chronological split is applied instead of random sampling because time-series forecasting must preserve the temporal order between past and future observations. This approach Is consistent with forecasting validation principles, where the testing data should represent future observations that are not used during model training. Previous studies also emphasize that validation design has a strong influence on forecasting performance, particularly when comparing fixed split and walk-forward validation techniques in ARIMA-based forecasting[12], [13]

E. Modelling

1. SARIMA Model

SARIMA (Seasonal Autoregressive Integrated Moving Average) is used as the first base learner. SARIMA extends the ARIMA model by incorporating seasonal components, making it suitable for time series data with recurring patterns. The model is generally represented as::

$$SARIMA(p,d,q)(P,D,Q)s$$

Where p,d,q represents the non-seasonal components of autoregressive, differencing, and moving-average orders, while P,D,Q represent the seasonal orders. The parameter s represents the seasonal period.

Previous studies indicate that SARIMA remains effective when temporal data can be transformed into a stationary structure and contain identifiable seasonal patterns [4], [5]. In this study, SARIMA parameter selection supported by stationarity testing, ACF and PACF analysis, and information criteria such as AIC and BIC. Manual parameter identification is also compared with the Auto-SARIMA approach to select the most suitable configuration.

2. Prophet Model

Prophet is applied as the second base learner. The model uses a decomposable additive structure consisting of trend, seasonality, holiday effects, and an error term [14]. Prophet is generally expressed as:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

Where *g(t)* represents the trend component, *s(t)* represents seasonality, *h(t)* represents holiday effects, and ϵ_t represents the error term

Prophet is included because its additive structure provides a complementary approach to SARIMA. While SARIMA focuses on autocorrelation and stationarity, Prophet models trend and seasonal behavior more explicitly. This distinction is relevant because empirical studies show that relative performance of SARIMA and Prophet depends on the characteristics of the observed time series [4], [5]..

3. XGBoost as Meta-learner

Academia Open
Vol. 11 No. 1 (2026): June
DOI: 10.21070/acopen.11.2026.14690

The prediction outputs from SARIMA and prophet are combined using a stacking ensemble approach. In this framework, both forecasting models act as base learners, while XGBoost acts as the meta-learner. The general formulation is expressed as:

$$\hat{y}_{stack} = f\left(\hat{y}_{SARIMA}, \hat{y}_{Prophet}\right)$$

Where $\hat{y}_{stack}$ represents the final ensemble prediction, while $\hat{y}_{SARIMA}$ and $\hat{y}_{Prophet}$ represent the prediction outputs generated by the two base learners.

Stacking is used because ensemble performance can improve when the meta-learner receives complementary information from both models with different forecasting structures [6], [15]. Jiang et al. also demonstrate that a meta-learning approach can improve prediction stability in short-term forecasting tasks with limited and fluctuating data [16]. In this study, XGBoost is selected because its tree-based boosting structure can model nonlinear relationships between base-model outputs while controlling complexity through regularization [17]

To reduce the risk of data leakage, the meta-learner is trained using out-of-fold predictions generated through time-series validation. In each fold, the base models are trained on earlier observations and used to predict the subsequent validation period. The resulting predictions are collected as meta-features, while the actual demand values are used as the targeted variable. After the meta-learner is trained, SARIMA and Prophet are refitted using the full training set and used to predict the testing period.

This validation strategy is used to ensure that the meta-learner is trained on predictions generated from temporally separated validation periods. Time-series validation is relevant for forecasting tasks because it prevents future observations from being used during model training and provides a more realistic evaluation structure than random cross-validation[13].

The meta-learner was configured using a relatively conservative parameter setting to reduce the risk of overfitting. The model employed 150 estimators with learning rate of 0.05. A maximum tree depth of 1 was selected to limit model complexity, while a subsample ratio of 0.6 was used so that each boosting iteration trained only on a portion of the available observations. The gamma parameter was set to 5 to restrict unnecessary tree splitting. In addition, regularization was strengthened by setting the L2 regularization parameter to 10 and the L1 to 5. These settings were applied to ensure that the meta-learner focused on stable correction patterns from the SARIMA and Prophet outputs rather than fitting short-term noise in the training data.

F. Evaluation

The evaluation phase compares the forecasting performance of SARIMA, Prophet, and the XGBoost-based stacking ensemble. Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are used as the primary evaluation metrics. RMSE measures the magnitude of prediction errors in the original unit, while MAPE expresses prediction errors as percentages [9].

RMSE is calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(A_t - \hat{A}_t)^2}$$

MAPE is calculated as:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - \hat{A}_t}{A_t}\right| \times 100\%$$

Where A_t represents the actual value, $\hat{A}_t$ represents the predicted value, and n represents the number of observations. Lower RMSE and MAPE values indicate better forecasting performance

III. Results and Discussion

A. Data Preparation and Stationarity Analysis

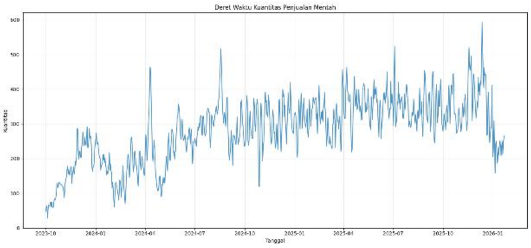
The preprocessing stage produced a continuous daily time series containing historical gallon demand data from October 1st, 2023, to January 22nd, 2026. The raw operational records were sorted chronologically, completed into a daily index, and processed using temporal interpolation to handle missing values. The historical visualization showed fluctuating daily demand with recurring weekly pattern as shown in Figure 2. Therefore, a seasonal period or 7 days to represent weekly seasonality in the SARIMA model.

Figure 2. Historical Daily Gallon Demand Time Series.

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Academia Open

Vol. 11 No. 1 (2026): June
DOI: 10.21070/acopen.11.2026.14690



Before model estimation, stationarity was evaluated using the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Philips-Schimdt-Shin (KPSS) tests. After first-order differencing, the ADF statistic reached -9.7485 with a p-value of 0.0000, while the KPSS reached 0.3337 with a p-value of 0.1000. These results indicate that the differenced series satisfies the stationarity requirement because the ADF test rejects the unit-root hypothesis and the KPSS test does not reject the stationarity hypothesis. Stationarity is important for SARIMA because the model estimates temporal relationships more effectively when the statistical structure of the series is sufficiently stable [9]. Detailed statistical output of this process is as shown i n Table 2 below..

Table 2. Stationarity Test After First-Order Differencing

Test	Statistic	p-value	Interpretation
ADF	-9.7485	0.0000	Stationary
KPSS	0.3337	0.1000	Stationary

The ACF and PACF plots were subsequently used to identify initial SARIMA parameter candidates. The PACF plot showed significant spikes at early lags, particularly lags 1, 2, 3, and 5, while the ACF plot showed a cut-off pattern after lag 3. These patterns indicate that the demand series contains short-term temporal dependence. However, parameter selection was not based exclusively on visual interpretation because several lags remained significant. Therefore, manual candidates were compared using AIC and BIC, and an automated parameter search was also conducted using Auto-SARIMA.

B. SARIMA Model Results

The manual SARIMA identification process produced several candidate configurations based on the ACF, PACF, and weekly seasonal period. Among the manually tested candidates, the selected configuration was SARIMA(3,1,1)(0,1,1)(7). In parallel, the Auto-SARIMA procedure identified SARIMA(2,1,3)(1,0,0)(7) as the most suitable automated configuration, with an AIC value of 6429.889 and a BIC value of 6461.4982.

The out-of-sample evaluation showed that the automated configuration performed better than the manually selected model. The manual SARIMA model produced an RMSE of 75.9720 and a MAPE of 19.42%, whereas Auto-SARIMA achieved an RMSE of 74.200307 and a MAPE of 18.49%. Since the final model selection was based on forecasting error on the testing period, Auto-SARIMA parameter was selected as the SARIMA base learner. The comparison between manual SARIMA and Auto-SARIMA is presented in Table 3 below.

Table 3. SARIMA Model Evaluation Results

Model	RMSE	MAPE
SARIMA(2,1,3)(1,0,0)(7)	75.9720	19.42%
SARIMA(3,1,1)(0,1,1)(7)	74.200307	18.49%

The stronger performance auto-SARIMA suggests that the demand series contains a temporal structure that can be effectively represented through autoregressive, moving-average, and weekly seasonal components. This finding is consistent with Gunawan and Ramadani [5], who reported that SARIMA outperformed Prophet in inventory forecasting when the seasonal structure of the data could be modeled more effectively. It also aligned with Kenyi and Yamamoto [18], who emphasized that SARIMA remains suitable when the series contains identifiable temporal dependence and can be transformed into a stationary structure.

C. Prophet Model Results

Prophet was developed as the second base learner to provide a forecasting structure that differs from SARIMA. The model was configured using weekly and yearly seasonality components to capture recurring temporal patterns. Unlike SARIMA, which relies on autocorrelation and differencing, Prophet decomposes the series into trend, seasonality, holiday effects, and residual components [14].

Table 4. Base Learner Evaluation Results

Base Learner	RMSE	MAPE
SARIMA	75.9720	18.42%
Prophet	92.954735	25.89%

The performance difference may indicate that the dominant information in the dataset is more closely related to short-term autocorrelation and weekly seasonality than to smoother trend changes. Prophet is designed to flexibly model trend and seasonal components, but this flexibility does not guarantee superior performance when daily fluctuations and short-term variations dominate the observed series. This interpretation is consistent with previous comparative studies showing that Prophet does not automatically outperform statistical models and that forecasting accuracy remains dependent on the characteristics of the

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

DOI: 10.21070/acopen.11.2026.14690

dataset [5], [18].

D. XGBoost-Based Stacking Ensemble Results

The SARIMA and Prophet predictions were subsequently used as meta-features for the XGBoost-based stacking ensemble. To avoid data leakage, the meta-training dataset was constructed using out-of-fold predictions generated through five-fold time-series validation. In each fold, the base learners were trained using earlier observations and evaluated on the subsequent validation period. The resulting SARIMA and Prophet predictions were then used as inputs for the XGBoost meta-learner.

The final stacking model achieved and RMSE of 73.648907 and a MAPE of 18.32%. These values are lower than the error values produced by both base learners. Therefore, the XGBoost-based stacking ensemble provides the best numerical performance among the evaluated models.

Table 5. Final Model Evaluation Results

Model	RMSE	MAPE
SARIMA	74.200307	18.49%
Prophet	92.954735	25.89%
Meta-Learner (XGBoost)	73.648907	18.32%

Compared with SARIMA as the strongest base learner, the stacking model reduced RMSE by 0.551400 and MAPE by 0.17%. These results show that the ensemble model improved forecasting accuracy, although the magnitude of improvement was relatively limited.

E. Discussion

The limited improvement obtained from the stacking model can be explained through the principle of base-learner diversity. An ensemble does not automatically produce a substantial performance improvement simply because several models are combined. Its effectiveness depends on whether the base learners provide complementary information and generate sufficiently different error patterns. When the base models capture different aspects of the data, the meta-learner has more useful information for correcting prediction errors [7], [19].

In this study, SARIMA and Prophet were selected because they represent different forecasting structures. SARIMA captures autoregressive and seasonal relationships, while Prophet models trend and seasonality through additive decomposition. However, the evaluation results show a clear performance imbalance: SARIMA produced a MAPE of 18.49%, whereas Prophet produced a MAPE of 25.89%. This difference may indicate that SARIMA already captured most of the dominant temporal information in the dataset, while Prophet provided only limited additional information for the meta-learner.

As a result, XGBoost was still able to learn correction patterns from the two base learners, but the available room for improvement was relatively narrow. The meta-learner reduced prediction error because it did not simply average the two outputs; instead, it learned how to adjust their contributions based on patterns observed during out-of-fold training. However, the weaker contribution from Prophet limited the potential improvement over the strongest base model.

This finding is important because it presents a more realistic interpretation of stacking ensemble performance. Jiang et al. [16] demonstrated that meta-learning can improve forecasting stability under fluctuating data conditions, while Wang and Dong [19] and Divina et al. [7] emphasized the importance of combining diverse learners. The present study supports these findings but also shows that ensemble improvement may remain marginal when the base learners do not contribute equally strong and complementary information.

From a practical perspective, the stacking ensemble is preferable when the primary objective is to minimize forecasting error. However, SARIMA remains a competitive alternative if implementation simplicity and lower computational complexity are prioritized. The main contribution of this study lies in demonstrating that the optimization process must be evaluated empirically rather than assumed to provide a large improvement in every forecasting context.

IV. Conclusions

This study applied SARIMA and Prophet as base learners and XGBoost as a meta-learner to forecast daily gallon demand at Depot XYZ. The results show that SARIMA achieved an RMSE of 74.200307 and a MAPE of 18.49%, while Prophet produced an RMSE of 92.954735 and a MAPE of 25.89%. The XGBoost-based stacking ensemble generated the best forecasting performance, with an RMSE of 73.648907 and a MAPE of 18.32%. Although the stacking ensemble reduced the prediction error compared with the strongest base learner, the improvement remained limited, with a reduction of 0.551400 in RMSE and 0.17 percentage points in MAPE. This finding indicates that the effectiveness of stacking depends not only on the use of multiple models but also on the diversity and complementary information provided by the base learners. In this case, SARIMA already captured most of the dominant temporal patterns in the dataset, while Prophet provided a more limited additional contribution to the meta-learner. Therefore, the XGBoost-based stacking ensemble can be selected as the final forecasting model when the primary objective is to minimize prediction error. However, SARIMA remains a competitive alternative when implementation simplicity and computational efficiency are prioritized. Future studies may incorporate external variables, such as holidays, weather conditions, or customer-related characteristics, to provide more diverse information and further improve forecasting accuracy.

References

1. Direktorat Jenderal Industri Kecil, Menengah, dan Aneka, Kementerian Perindustrian Republik Indonesia, "Kolaborasi dengan Kemenkes, Kemenperin Gelar Kursus Tingkatkan Standar Sanitasi Depot Air Minum." Accessed: Jun. 04, 2026. [Online].
2. Badan Pusat Statistk, "Persentase Rumah Tangga yang Memiliki Akses terhadap Sumber Air Minum Layak Menurut Provinsi (Persen), 2025," bps.go.id. Accessed: Jan. 29, 2026. [Online]. Available:
3. A. Razak Thaha and la Al Maududi, "Enrichment: Journal of Multidisciplinary Research and Development Analysis of the Relationship Between Drug Stockout Incidents and the Use of Hospital Management Information System (SIMRS) Technology with Patient Satisfaction at the Pharmacy Installation of Bogor City Hospital," 2025. doi: https://doi.org/10.55324/enrichment.v3i6.442.
4. W. Chango et al., "SARIMA vs. Prophet: Comparative Efficacy in Forecasting Traffic Accidents Across Ecuadorian Provinces," Computation, vol. 14, no. 1, p. 5, 2025, doi: https://doi.org/10.3390/computation14010005.
5. W. Gunawan and M. Ramadani, "Analisa Perbandingan Penerapan Metode SARIMA dan Prophet dalam Memprediksi Persediaan Barang PT XYZ," Faktor Exacta, vol. 16, no. 2, pp. 88–97, 2023, doi: https://doi.org/10.30998/faktorexacta.v16i2.13803.
6. A. Seyam, S. S. Mathew, B. Du, M. El Barachi, and J. Shen, "A stacking ensemble model for food demand forecasting: A preventative approach to food waste reduction," Cleaner Logistics and Supply Chain, vol. 15, Jun. 2025, doi: https://doi.org/10.1016/j.clscn.2025.100225.
7. F. Divina, A. Gilson, F. Goméz-Vela, M. G. Torres, and J. F. Torres, "Stacking ensemble learning for short-term electricity consumption forecasting," Energies (Basel)., vol. 11, no. 4, pp. 1–31, doi: https://doi.org/10.3390/en11040949.
8. R. J. Hyndman and G. Athanasopaulos, Forecasting: Principles and Practice, 2nd ed., 2018. [Online].
9. I. G. T. Suryawan, I. K. N. Putra, P. M. Meliana, and I. G. I. Sudipa, "Performance Comparison of ARIMA, LSTM, and Prophet Methods in Sales Forecasting," Sinkron, vol. 8, no. 4, pp. 2410–2421, 2024, doi: https://doi.org/10.33395/sinkron.v8i4.14057.
10. C. Pete et al., "Crisp-Dm 1.0," CRISP-DM Consortium, p. 76, 2000.
11. Y. Liu, D. Kalaitzi, M. Wang, and C. Papanagnou, "A machine learning approach to inventory stockout prediction," Journal of Digital Economy, vol. 4, pp. 144–155, Jan. 2025, doi: https://doi.org/10.1016/j.jdec.2025.06.002.
12. D. Arya Inasta Asrul and A. Andy Soebroto, "Optimalisasi Prediksi Kasus COVID-19 di Indonesia: Perbandingan Teknik Validasi 80-20 Split dan Walk-Forward dengan ARIMA," J-INTECH, vol. 12, no. 02, pp. 297–308, 2022, doi: https://doi.org/10.32664/j-intech.v12i02.1373.
13. W. Sulandari, Y. Yudhanto, S. Subanti, E. Zukhronah, and M. Z. Subarkah, "Implementing Time Series Cross Validation to Evaluate the Forecasting Model Performance," KnE Life Sciences, Mar. 2024, doi: https://doi.org/10.18502/kls.v8i1.15584.
14. S. J. Taylor and B. Letham, "Forecasting at Scale", doi: https://doi.org/10.7287/peerj.preprints.3190v2.
15. X. Wang, R. J. Hyndman, F. Li, and Y. Kang, "Forecast combinations: an over 50-year review," Int. J. Forecast., vol. 39, no. 4, pp. 1518–1547, Sep. 2022, doi: https://doi.org/10.1016/j.ijforecast.2022.11.005.
16. G. Jiang, G. Han, M. Zhao, and W. Ma, "Meta-Learning Enhanced Random Forest Ensemble for Short-Term Oil Well Production Forecasting," Processes, vol. 13, no. 12, pp. 1–18, 2025, doi: https://doi.org/10.3390/pr13123797.
17. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, vol. 13-17-August-2016, pp. 785–794, Mar. 2016, doi: https://doi.org/10.1145/2939672.2939785.
18. M. G. S. Kenyi and K. Yamamoto, "A hybrid SARIMA Prophet model for predicting historical streamflow time-series of the Sobat River in South Sudan," Discover Applied Sciences, 2024, doi: 10.1007/s42452-024-06083-x.
19. Y. Wang and Y. Dong, "A Selective Ensemble Learning Method for Time Series Forecasting," MLNLP 2024 - 2024 International Conference on Machine Learning and Natural Language Processing, 2024, doi: https://doi.org/10.1109/MLNLP63328.2024.10800007.

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo