



Demand Forecasting for Industrial Gas Distribution in Small and Medium Enterprises: Evaluating Model Suitability under Different Demand Characteristics

Rizky Zaqi Megantara¹, Mewati Ayub^{2*}, Oscar Karnalim³

Master of Computer Science, Universitas Kristen Maranatha, Bandung 40164, Indonesia

Corresponding Author Email: mewati.ayub@it.maranatha.edu

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.310922>

ABSTRACT

Received: 14 May 2026

Revised: 20 August 2026

Accepted: 9 September 2026

Available online: 30 September 2026

Keywords:

demand forecasting, industrial gas distribution, time-series forecasting, forecasting model selection, demand variability, small and medium enterprises, machine learning

Accurate demand forecasting is essential for inventory planning and operational decision-making in small and medium-sized enterprises (SMEs), particularly in industrial gas distribution where demand patterns are often volatile and historical data availability is limited. However, selecting an appropriate forecasting approach remains challenging because model performance is highly dependent on the underlying characteristics of demand data. This study investigates the suitability of statistical and machine learning forecasting methods for industrial gas demand prediction in an SME distribution environment. Three representative forecasting approaches, including Prophet, seasonal autoregressive integrated moving average (SARIMA), and Extreme Gradient Boosting (XGBoost), were evaluated using weekly transaction data collected from five major industrial gas products between 2022 and 2024. The forecasting performance was assessed using mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) under multiple temporal validation scenarios. The results demonstrate that no single forecasting model consistently outperforms other approaches across different demand profiles. Prophet achieved reliable performance for relatively stable products, while simpler baseline methods remained competitive for highly volatile products with weak temporal structures. These findings indicate that increasing model complexity does not necessarily lead to improved forecasting accuracy. Instead, forecasting model selection should be guided by demand variability, temporal characteristics, and practical operational requirements. This study provides empirical insights into demand forecasting strategies for SME industrial gas distributors facing heterogeneous product demand patterns.

1. INTRODUCTION

Demand forecasting is the process of predicting future customer needs for a product or service, enabling companies to optimally adjust procurement strategies and inventory management [1]. In practice, fluctuating sales and potential customer loss are crucial issues for business operators, especially when companies operate under intense competition with limited resources. Therefore, an accurate demand forecasting system is essential to support efficient, data-driven operational decision-making [2].

This problem is particularly relevant in the industrial gas industry, where companies must ensure the availability of essential gas supplies such as oxygen, nitrogen, and argon. According to a report from Research and Markets [3], Indonesia's industrial gas market is experiencing significant growth from USD 119.90 billion in 2024 to USD 128.28 billion in 2025. This growth is projected to continue with a Compound Annual Growth Rate of 7.30% to reach USD 183 billion by 2030.

In the industrial gas distribution sector, small and medium-sized enterprises (SMEs) typically procure gas from suppliers on a weekly basis without systematic demand estimation

support. This practice frequently leads to inventory imbalances, as over-ordering drives up storage costs while under-ordering risks stockouts and lost revenue.

Time-series analysis is one of the most common approaches to demand forecasting, especially in retail, due to its ability to capture trends and seasonal behavior [4]. Traditional methods based on probability and statistics have proven their superiority in various fields such as meteorology, finance, and industry. However, with the development of the big data era, increasingly massive, non-linear time-series data following various distribution patterns have emerged, presenting new challenges for forecasting methods. Therefore, machine learning and deep learning have begun to be applied to predict highly complex time series and have shown excellent results [5].

This research focuses on evaluating the performance of Prophet, seasonal autoregressive integrated moving average (SARIMA), and Extreme Gradient Boosting (XGBoost) methods in improving demand forecasting accuracy in the SME context. SARIMA was chosen for its ability to capture temporal dependencies and seasonal patterns in time-series data [5]. XGBoost was selected based on its flexibility in handling data with weak autocorrelation and high variability.

Unlike traditional time-series models like SARIMA, which heavily depend on strong temporal dependencies, XGBoost uses a tabular feature approach capable of extracting patterns from exogenous or contextual variables [6]. Meanwhile, Prophet was chosen for its ability to flexibly model trends and seasonal components, as well as its adaptability to pattern changes in time-series data, including calendar influences and special events [7].

The performance of each model was evaluated using mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) metrics commonly used in demand forecasting evaluation. MAE measures average absolute error levels, RMSE assesses model sensitivity to large errors (outliers), while MAPE represents relative error to data scale. Besides these error-based metrics, nominal 95% forecast intervals were used only as qualitative uncertainty diagnostics; model ranking was based on MAE, RMSE, and MAPE.

For the industrial gas distribution section, accurate forecasting is important to keep daily operations running smoothly. Industries can prepare enough gas based on the forecasted pattern. It prevents the costly waste of liquid gas evaporating into thin air from overstocked storage tanks.

2. LITERATURE REVIEW

2.1 Forecasting methods: Statistical vs. machine learning

The debate between traditional statistical methods and machine learning approaches in time-series forecasting has been extensively studied. Cerqueira et al. [8] conducted a comprehensive study which showed that statistical methods consistently outperformed machine learning on small datasets. Their research using 90 time series with a minimum of 1000 observations revealed that while statistical methods excel on small datasets, machine learning methods show better performance as data size increases. This finding emphasizes the importance of considering dataset size when selecting forecasting approaches.

In the SME manufacturing context, research by Aidoo-Anderson et al. [9] examining demand forecasting practices in pharmaceutical SMEs showed that most companies still rely on judgmental forecasting and simple methods due to limited historical data, human resources, and low technology adoption. The study confirmed that SME demand is often volatile, sporadic, and exhibits intermittent demand characteristics, making stable historical patterns difficult to identify.

A benchmark study across industries by Syberg et al. [10] comparing statistical and machine learning methods in make-to-order production environments characterized by highly fluctuating demand and limited historical data found that no single method consistently outperformed others under all conditions. Simple statistical methods such as ARIMA, Seasonal Naïve, and exponential smoothing often provided more stable performance compared to machine learning models.

2.2 Extreme Gradient Boosting in forecasting applications

XGBoost has emerged as a powerful tool for forecasting tasks due to its ensemble tree-based approach. Martinović et al. [11], in their study on predicting innovation outcomes at the company level, compared various machine learning

algorithms, including Random Forest, XGBoost, CatBoost, SVM, and ANN. Results showed that boosted decision tree models, particularly XGBoost, excelled in accuracy, F1-score, and ROC-AUC metrics. The study emphasized the importance of using corrected cross-validation techniques and Bayesian-based hyperparameter optimization.

Acar and Gardner [6] conducted a case study of forecasting method selection for a global manufacturer, comparing exponential smoothing methods against optimization and simulation techniques to minimize total supply chain cost. Their results showed that the choice of forecasting method has a direct, measurable impact on downstream supply chain cost and service level, reinforcing that method selection should be driven by empirical performance on the specific demand pattern rather than by model sophistication alone.

2.3 Prophet for business forecasting

Facebook Prophet has gained widespread adoption in business forecasting due to its ability to model long-term trends, seasonality, and calendar/holiday effects within a relatively easy-to-implement additive model framework. Prophet's main advantage lies in its flexibility in handling trend changes (change points), multi-seasonality, and robustness to non-stationary data commonly found in industrial and SME sales data.

Žunić et al. [12] developed a retail sales forecasting framework based on Prophet, equipped with rolling back-testing procedures to evaluate forecast reliability at the product (SKU) level. The study showed that Prophet can provide stable performance across various forecasting horizons and offers good interpretability through trend and seasonal component decomposition.

Petropoulos et al. [13] evaluated forecasting method performance across thousands of real time series and found that no single method, however sophisticated, consistently outperforms others across all data patterns. Their findings support a "horses for courses" approach in which the choice between methods such as Prophet, SARIMA, or simple exponential smoothing should be guided by the specific characteristics of the data at hand, such as volatility and trend or seasonal strength, rather than by a default preference for model complexity.

Furthermore, Prophet has been combined with other algorithms in hybrid approaches. Guo et al. [14] showed that combining Prophet with Support Vector Regression (SVR) can improve forecasting accuracy by leveraging Prophet's strength in capturing trends and seasonality, along with machine learning's capability in modelling non-linear relationships not captured by Prophet's additive structure.

2.4 Seasonal autoregressive integrated moving average in industrial applications

SARIMA remains one of the most widely used statistical methods for time-series forecasting, particularly for data with strong seasonal patterns. The model's ability to capture both short-term dependencies and long-term seasonal cycles makes it suitable for various industrial applications.

Falatouri et al. [4] compared SARIMA and LSTM in retail supply chain management, finding that SARIMA provided competitive performance with easier interpretability and lower computational requirements compared to deep learning approaches. The study emphasized SARIMA's effectiveness

in capturing regular seasonal patterns without requiring extensive computational resources.

However, SARIMA's performance depends heavily on data characteristics. The model requires sufficient historical data to estimate parameters accurately and assumes linear relationships between variables. In contexts with high noise levels or irregular patterns, simpler methods may outperform SARIMA despite its theoretical sophistication.

2.5 Industrial gas forecasting and B2B demand prediction

Giannopoulos et al. [15] reviewed machine learning approaches for intermittent and lumpy demand forecasting, covering hyperparameter tuning, feature engineering, and training strategies across 124 peer-reviewed studies. Their review highlighted that ML methods for intermittent demand require careful handling of overfitting and the risk of treating noise as signal on short, volatile series. Zhou et al. [16] presented a predictor with denoising features and information combination to guide contrastive representation learning for gas forecasting. Sanchez-Ubeda and Berzosa [17] captured the trend and combined it with the Linear Hinges Model and a transitory component to predict gas consumption.

Qin et al. [18] addressed data scarcity through a Federated Contrastive Pretraining-Local Clustered Fine-tuning paradigm. Pu et al. [19] used a spatial-temporal Bayesian model to predict gas demand through time-varying spatial graphs, considering gas-pipeline connectivity and co-movement. Syntetos et al. [20] critically reviewed the gap between forecasting theory and supply chain practice, noting that academic research has often neglected practical issues such as

data quality and the operational translation of theoretical forecasting gains.

2.6 Research gap

While extensive research exists on comparing forecasting methods, few studies specifically address the SME industrial gas distribution context characterized by: (1) limited historical data availability, (2) high demand volatility without clear seasonal patterns, (3) B2B customer relationships creating dependency on client business activities, and (4) resource constraints limiting complex model implementation.

This research contributes by evaluating how different forecasting approaches perform across products with varying demand characteristics (measured by coefficient of variation (CV)) in a real SME setting, providing practical insights for similar businesses facing comparable challenges.

3. METHODOLOGY

Figure 1 illustrates the overall research flow applied in this study. The process begins with raw transaction data collection and preprocessing, followed by exploratory data analysis (EDA) to understand demand characteristics. Three forecasting models—Prophet, SARIMA, and XGBoost—are then developed and evaluated against baseline benchmarks. Model performance is assessed using multiple error metrics and experimental train-test splits to ensure robustness across different temporal conditions.



Figure 1. Research flow

3.1 Data collection and preprocessing

This research uses historical transaction data from an industrial gas distribution SME's database system, covering daily industrial gas sales from 2009–2024. Based on trend analysis, the 2022–2024 period was selected as the study basis due to data stability and operational relevance. Years prior to 2022 showed structural inconsistencies; in particular, 2018 was excluded due to systematic missing data caused by high employee turnover. The final study dataset spans three years (2022–2024) and contains 76,200 raw transactions across all products. The dataset represents a typical B2B demand environment, with 96% of transactions occurring on weekdays, reflecting the industrial nature of the customer base. Each transaction acts as a data point with the product name, issue week (calculated from the issue date), and the number of cylinders sold.

Product selection was limited to the top five gas products based on transaction frequency: Oxygen, Argon, Nitrogen, Acetylene, and Argon UHP, collectively covering 89.9% of total transactions (approximately 68,500 out of 76,200 total transactions). Table 1 provides an overview of the demand

characteristics of these five products. Daily transaction data were aggregated into weekly frequency to reduce daily noise and capture clearer seasonal patterns, resulting in approximately 157 weekly observations per product for the 2022–2024 study period.

Table 1. Descriptive statistics for top 5 products (weekly data)

Product	Mean	Median	Standard Deviation	Coefficient of Variation (CV) (%)
Oxygen	297.71	300.0	72.32	24.29
Argon	48.12	48.5	14.68	30.50
Nitrogen	27.81	26.0	12.07	43.49
Acetylene	24.99	22.0	12.27	49.10
Argon UHP	11.47	9.5	8.47	73.80

3.2 Exploratory data analysis

Before modelling, in-depth data analysis was conducted to understand data characteristics and patterns. The EDA stages

included:

- Descriptive statistics: Calculation of mean, median, standard deviation, and CV for each product at weekly aggregation.

- Autocorrelation analysis (ACF and PACF): Analysis of temporal dependency structure on weekly sales data to identify relationships between current sales values and previous periods (lags). Analysis was conducted up to 52 lags to detect both short-term dependencies and annual seasonal cycles.

- Time series decomposition: Data decomposition using the seasonal and trend decomposition using loess (STL) method to separate data into three main components: trend-cycle, seasonal, trend strength (F_T), and seasonal strength (F_S). The calculations were performed to provide a deeper descriptive understanding of STL plot results using variance-based formulas [21].

3.3 Forecasting models

This research compared the performance of three algorithms representing different approaches in time-series forecasting:

- Prophet: This method was chosen for its ability to handle non-linear trends and strong seasonal effects with an approach robust to missing values and outliers. The model configuration used additive seasonality mode with yearly, weekly, and daily seasonality components, a 95% confidence interval width, and a changepoint prior scale.

- SARIMA: Optimal parameters were determined using the auto-ARIMA algorithm with a stepwise search approach. The best model selection criterion was based on the lowest Akaike Information Criterion (AIC) value.

- XGBoost: As a representation of non-linear ensemble tree-based machine learning methods, XGBoost requires time-series data transformation into supervised learning format. The feature engineering process constructs explicit temporal signals, covering ten features. Autoregressive lag features are at 1, 4, 13, and 52 weeks. Rolling-window statistics smooth short-term fluctuations and measure changing volatility using strictly past observations to eliminate data leakage. To represent calendar cycles without introducing artificial discontinuities between consecutive periods, calendar features are transformed using sine and cosine functions into continuous two-dimensional coordinates. Finally, relative momentum features, including short-term differences and year-over-year ratios, explicitly isolate velocity and growth trajectories across various time horizons. The hyperparameter tuning resulted in $n_estimators = 1000$, $learning_rate = 0.025$, and $max_depth = 4$. Autoregressive lags were the most important ones.

Baseline models (Average, Naïve, and Seasonal Naïve) were also included as benchmarks to evaluate whether complex models provided significant improvement over simple statistical approaches.

3.4 Evaluation metrics

To measure forecasting error levels on test data, this research uses three statistical indicators as follows.

MAE measures the average magnitude of absolute prediction errors in the original transaction unit (cylinders per week), providing an interpretable baseline error level.

RMSE penalizes large errors more heavily than MAE, making it sensitive to outlier predictions — a useful signal for

identifying models that fail during irregular demand periods.

MAPE expresses the error as a percentage of the actual value, enabling scale-independent comparisons across products with very different demand volumes.

For methods that produced interval estimates, nominal 95% forecast intervals were plotted and inspected qualitatively, with emphasis on their width across products and forecast horizons. No empirical coverage statistic or calibration/reliability test was calculated, and interval estimates were not used to rank the models. The quantitative comparison therefore relies on MAE, RMSE, and MAPE.

Combining these metrics provided a balanced overview of the model performance, especially for highly volatile products. Volatile demand combines sharp demand spikes, rapid trend shifts, and high residual noise. No single metric captures both average operational performance and extreme risk.

3.5 Experimental design

Data splitting was conducted to test model consistency and stability under various temporal conditions. The dataset used for all experiments consists of weekly aggregated transaction data for five industrial gas products—Oxygen, Argon, Nitrogen, Acetylene, and Argon UHP—spanning the period January 2022 to December 2024 (157 weekly observations per product). The data is quantitative and transactional (number of cylinders sold per week), derived from a B2B distribution context in which 96% of demand occurs on weekdays.

There were four experimental scenarios, assuming that some periods might influence prediction accuracy. The first experiment considered the full year as one cycle. Data from 2024 acted as the testing data, while data from 2022 and 2023 acted as the training data for building the model. This showed our model performance while considering all available data. The second experiment was similar to the first one except that the analysis was conducted per quarter. This narrowed down the analysis. The models might be more effective in one quarter than another. The third experiment considered data from 2023 for testing and data from 2022 for training. The fourth experiment considered data from 2024 for testing and data from 2023 for training. The last two experiments examined whether there were anomalies between 2022 and 2024 due to the COVID pandemic. The analysis can be seen in Section 4.3.

4. RESULTS AND DISCUSSION

4.1 Data characteristics

Based on descriptive statistical analysis, five main products showed different variability levels reflected in their CV values. Table 1 presents descriptive statistics for the five products based on weekly aggregation.

Oxygen shows the highest transaction volume, with 44,553 transactions (approximately 65% of the combined total for the top five products, or 58.5% of the full 76,200-transaction dataset), with an average of 284.64 transactions per week. Besides high volume, this product shows relatively good demand stability, reflected in the lowest CV value of 24.29% among the five analyzed products.

Conversely, Argon UHP shows the highest volatility with CV reaching 73.80%, although transaction volume is relatively small at 1,987 transactions or approximately 2.9%

of total transactions. High volatility is closely related to Argon UHP usage characteristics, as high-purity industrial gas is generally used in laboratory testing activities and precision analytical instrument operation.

Time series decomposition results using the STL method showed that all gases have Seasonal Strength significantly more dominant than Trend Strength. Trend Strength values for all gases are in the low to very low category (below 0.2), while Seasonal Strength values range between 0.40 and 0.48. The seasonal strength value is likely influenced by the Eid al-Fitr holiday cycle, which, as a moving holiday, causes a total operational shutdown, recorded as a drastic sales volume reduction to zero in the observed data.

ACF and PACF analysis results showed that the weekly industrial gas demand time series exhibits relatively weak autocorrelation characteristics. Most ACF and PACF values at various lags are within statistical confidence bounds, with only a few small spikes appearing sporadically. This condition indicates that the relationship between current demand values and previous period values does not show strong and systematic temporal dependency.

In summary, for low-volatility products like Oxygen and Argon, demand was stable and symmetrical, as indicated by near-identical mean and median values. SARIMA was the optimal model for these two because its linear and parametric structure relied on strong autocorrelation and Gaussian residual distributions.

As volatility increases in Nitrogen and Acetylene, the widening gap between mean and median values indicated right-skewed demand. For these products, Prophet served as the best solution due to its additive piecewise linear regression, which automatically detected trend change points and multi-period seasonality. It also remains resilient to demand spikes.

For highly volatile and lumpy demand profiles like Argon UHP, neither the statistical time-series models nor XGBoost were effective. In this scenario, the simple Mean baseline was actually the best-performing approach (MAPE 57.66%, Table 2), outperforming XGBoost, SARIMA, and Prophet. This indicates that when demand approaches white noise with no learnable structure, added model complexity provides no benefit and can instead amplify overfitting to spurious patterns.

4.2 Model performance

Table 2 presents a summary of the best model performance for each product. For Oxygen, the Prophet model showed the best performance with a MAPE of 16.82% when using 2022-2023 training data to predict 2024. However, quarterly evaluation revealed significant performance degradation in Q2 (MAPE increased to 30.82%) due to the Eid al-Fitr holiday shift, which the model without external holiday variables could not anticipate. Prophet's relatively good performance is supported by low data fluctuation levels (CV 24.29%), indicating sales tend to be around their average value. Rolling forecast evaluation with weekly retraining showed Prophet achieved a MAPE of 15.05%, demonstrating strong performance for short-term operational predictions when using the most recent data patterns.

To address all four experimental scenarios described in Section 3.5, Table 3 consolidates the training period, testing period, and best-performing model (lowest MAPE) for each product across the three fixed-window train-test splits. The fourth scenario (quarterly granularity) is summarized narratively above and in Section 4.3, since its results are reported per quarter rather than as a single best model.

Table 2. Summary of the best model performance for each product

Product	Best Model	MAE	RMSE	MAPE (%)	CV (%)
Oxygen	Prophet	44.94	61.72	16.82	24.29
Argon	Mean	10.79	13.34	24.85	30.50
Nitrogen	SARIMA	6.34	8.26	28.48	43.49
Acetylene	Mean	9.67	11.57	96.85	49.10
Argon UHP	Mean	7.16	10.01	57.66	73.80

Note: SARIMA = seasonal autoregressive integrated moving average, MAE = mean absolute error, RMSE = root mean squared error, MAPE = mean absolute percentage error, CV = coefficient of variation.

Table 3. Summary of best-performing model across the three fixed train-test scenarios (Section 3.5, Scenarios 1–3)

Scenario	Training Period	Testing Period	Product	Best Model	MAPE (%)
Scenario 1	2022–2023	2024	Oxygen	Prophet	18.77
			Argon	SARIMA	23.27
			Nitrogen	SARIMA	28.48
			Acetylene	XGBoost	84.34
			Argon UHP	Mean	57.66
Scenario 2	2023	2024	Oxygen	Prophet	17.06
			Argon	Naïve	24.75
			Nitrogen	SARIMA	27.21
			Acetylene	Naïve	46.09
			Argon UHP	Mean	62.52
Scenario 3	2022	2023	Oxygen	Naïve	19.50
			Argon	Prophet	24.96
			Nitrogen	XGBoost	30.98
			Acetylene	Prophet	47.43
			Argon UHP	XGBoost	78.95

Note: SARIMA = seasonal autoregressive integrated moving average, XGBoost = Extreme Gradient Boosting, MAPE = mean absolute percentage error.

Across all three scenarios, the best-performing model varies by product and by training window: Oxygen is consistently

best served by Prophet or, when only one year of recent data is available, by the Naïve baseline; Argon and Nitrogen

alternate between SARIMA, Naïve, XGBoost, and Prophet depending on the period; and Acetylene and Argon UHP, the two highest-CV products, are best served by simple baselines (Mean, Naïve) or, in one scenario, XGBoost, but never consistently by a single complex model. This instability across scenarios reinforces that no single algorithm generalizes across training windows for this dataset, and that reported “best models” should be read as scenario-specific rather than as a fixed model recommendation per product.

For Argon and Nitrogen: with CV values around 30% and 40% respectively, machine learning models did not show significant forecasting performance improvement compared to baseline methods. Simple benchmark approaches such as Naïve forecasting were able to produce accuracy levels comparable to more complex models. This indicates that although demand variability levels are relatively high, fluctuations are not accompanied by consistent patterns learnable by complex models. For Argon, the best-performing model (Mean method) achieved a MAPE of 24.85% and MAE of 10.79 cylinders per week. Nitrogen followed a similar pattern, with SARIMA marginally outperforming other approaches at MAPE 28.48%, though the improvement over simpler methods was modest. The relatively similar performance across models for both products can be attributed to the nature of their customer base: these gases serve a wider variety of industries and use cases than Oxygen, resulting in demand that is more fragmented and harder to predict from historical patterns alone. Quarterly analysis further revealed that Q2 performance degraded for all models across both products, consistent with the Eid al-Fitr disruption observed in Oxygen.

For Acetylene and Argon UHP: with the highest CV values of 52% and 70%, respectively, both products present difficulties for both machine learning models and baseline methods in producing accurate forecasts. This is reflected in high MAPE values, ranging from above 50% to even exceeding 100% in most forecasting models. High fluctuation levels in Acetylene products are caused by customer company characteristics, where most customers do not use Acetylene as the main material in production processes, making demand uncertain and without consistent patterns. For Argon UHP, extreme volatility is driven by its role as a specialty gas primarily used in laboratory and precision analysis contexts, where purchasing is project-driven rather than routine, meaning demand spikes are largely unpredictable without external context. In both cases, even the best-performing model (Mean method) produced MAPE values of 96.85% for Acetylene and 57.66% for Argon UHP, indicating that these products may be better managed through safety stock policies or order-point systems rather than traditional time-series forecasting. These results align with findings in the broader literature that products with near-white-noise demand patterns are better served by inventory buffering strategies than by predictive modelling.

4.3 Impact of training window

Experiments with various training data durations, namely two years and one year, showed that using two years of training data did not produce significant forecasting performance improvement compared to using one year of training data. This finding indicates that forecasting based on more recent historical data provides a more relevant representation of actual demand conditions.

For example, in Oxygen product forecasting, Prophet training using only 2023 data produced an MAE of 44.90 and a MAPE of 17.06%, almost identical to training results using two years of data (2022–2023). Very small performance differences show that data recency has a more dominant role than historical horizon length.

Taken together, the cross-product results reveal a clear relationship between demand predictability and model suitability. As illustrated in Table 2, MAPE values increase substantially as the CV of the product rises, from 16.82% for Oxygen (CV 24.29%) to 96.85% for Acetylene (CV 49.10%). This gradient underscores that no single model consistently dominates; rather, forecasting strategy should be tailored to each product’s demand structure. From a practical standpoint, this finding has direct operational implications. SME industrial gas distributors with a mixed product portfolio comparable to the one studied here would benefit from a tiered forecasting approach: deploying statistical models such as Prophet for stable, high-volume products, while relying on simpler safety-stock policies for high-volatility items. Furthermore, the consistent Q2 performance degradation across all models and products highlights that model-agnostic external information, such as holiday calendars or B2B customer activity schedules, could substantially improve forecast reliability during disruption-prone periods.

Qualitative inspection of the forecast plots showed generally wide nominal 95% forecast intervals, particularly for more volatile products and longer forecast horizons. This pattern indicates substantial uncertainty and limited forecast precision, but interval width alone cannot establish whether nominal 95% coverage was achieved. Formal evaluation of interval coverage and sharpness is therefore outside the scope of this study.

4.4 Key findings

Several important findings from the study can be obtained. Model complexity does not always guarantee better accuracy. On products with high fluctuation levels and weak time-series structure or approaching white noise, complex models like SARIMA and XGBoost did not show significant superiority over baseline methods. In such conditions, simple statistical methods like Mean and Naïve provided more stable and competitive forecasting results.

When time-series data exhibits high fluctuation and weak temporal structure, simple methods exceed complex models because they avoid overfitting to random noise and directly align with underlying stochastic processes. High-capacity architectures like SARIMA and XGBoost require stable signals to estimate autoregressive parameters or feature split thresholds. By assuming minimal structural form, these baseline methods eliminate parameter estimation instability and maintain robust predictive performance.

CV shows its importance as a predictability indicator. The higher a product's CV value, the more difficult it is for forecasting models to produce accurate predictions. This is caused by the company's data characteristics: high CV values are not driven by strong, recurring seasonal patterns, but rather by the data's fluctuating and unstructured nature.

Data recency is important. Using more recent training data provides a better representation of current market conditions compared to longer historical data, which may contain patterns that are no longer relevant.

There is a need for external variables. Significant

performance degradation in Q2 across all models emphasizes the need to integrate external variables such as national holiday calendars or holiday regressors to improve prediction accuracy, especially during religious holiday periods.

5. CONCLUSIONS

Within the scope of this single-SME case study, this research concludes that industrial gas demand characteristics are strongly influenced by each product's variability level. The five main products consisting of Oxygen, Argon, Nitrogen, Acetylene, and Argon UHP cover 89.9% of total transactions, with Oxygen as the most dominant product having the highest demand stability level.

From a forecasting model performance perspective, the results show that algorithm complexity is not always directly proportional to improvements in prediction accuracy. Prophet model proved most reliable in forecasting relatively stable Oxygen demand, with a MAPE value of 16.82%. However, for products with high demand fluctuation and weak time-series structure or approaching white noise, complex models like SARIMA and XGBoost did not show significant superiority over baseline methods.

This research also confirms that temporal data relevance or historical data recency has greater influence on forecasting accuracy compared to training data period length. Use of the most recent one-year training data proved capable of better representing current market conditions compared to using two years of data potentially containing no longer relevant demand patterns.

Based on these findings, management at the studied company and similarly structured SME industrial gas distributors can apply forecasting strategies tailored to each product's characteristics; broader generalization beyond this single-SME context should be validated with additional case studies. More sophisticated forecasting models, such as Prophet, can be used for products with relatively stable demand patterns, while simple average-based methods can serve as baselines for products with high demand fluctuation levels. For future research, it is recommended that forecasting models integrate external variables, such as national holiday calendars or holiday regressors, to improve prediction accuracy, especially during holiday periods.

For future work, we have several directions. First, we plan to integrate more features, such as news trends and external holiday seasons. Second, we are also interested in employing feature selection to determine which features are more important than others. Third, it would be useful to address potential overfitting issues through additional evaluation metrics. Fourth, while our evaluation metrics are acceptable, we believe that other metrics might improve the quality of the findings. Using symmetric mean absolute percentage error (sMAPE) or mean absolute scaled error (MASE) might complement our metrics. Future work should also formally assess prediction-interval calibration using empirical coverage and interval-width or sharpness measures. Fifth, we want to address overfitting, which may occur in the models due to the dataset's characteristics.

REFERENCES

[1] Chowdhury, A.R., Paul, R., Rozony, F.Z. (2025). A

systematic review of demand forecasting models for retail e-commerce enhancing accuracy in inventory and delivery planning. *International Journal of Scientific Interdisciplinary Research*, 6(1): 1-27. <https://doi.org/10.63125/mbbfw637>

[2] Ul Haq Qureshi, N., Javed, S., Javed, K., Meesam Raza Naqvi, S., Raza, A., Saeed, Z. (2024). Demand forecasting in supply chain management for Rossmann stores using weather enhanced deep learning model. *IEEE Access*, 12: 145570-145581. <https://doi.org/10.1109/ACCESS.2024.3472499>

[3] Industrial Gases Market by Type (Oxygen, Nitrogen, Hydrogen, Carbon Dioxide, Acetylene, Inert Gases), End-use Industry (Chemical, Electronics, Food & Beverage, Healthcare, Manufacturing, Metallurgy, Refining), and Region - Global Forecast to 2030. <https://www.researchandmarkets.com/reports/5940261/industrial-gases-market-type-oxygen-nitrogen#src-pos-8>, accessed on Feb. 8, 2026.

[4] Falatouri, T., Darbanian, F., Brandtner, P., Udokwu, C. (2022). Predictive analytics for demand forecasting: A comparison of SARIMA and LSTM in retail SCM. *Procedia Computer Science*, 200: 993-1003. <https://doi.org/10.1016/J.PROCS.2022.01.298>

[5] Liu, Z., Zhu, Z., Gao, J., Xu, C. (2021). Forecast methods for time series data: A survey. *IEEE Access*, 9: 91896-91912. <https://doi.org/10.1109/ACCESS.2021.3091162>

[6] Acar, Y., Gardner Jr, E.S. (2012). Forecasting method selection in a global supply chain. *International Journal of Forecasting*, 28(4): 842-848. <https://doi.org/10.1016/j.ijforecast.2011.11.003>

[7] Stefenon, S.F., Seman, L.O., Mariani, V.C., Coelho, L.dosS. (2023). Aggregating prophet and seasonal trend decomposition for time series forecasting of Italian electricity spot prices. *Energies*, 16(3): 1371. <https://doi.org/10.3390/EN16031371>

[8] Cerqueira, V., Torgo, L., Soares, C. (2022). A case study comparing machine learning with statistical methods for time series forecasting: Size matters. *Journal of Intelligent Information Systems*, 59(2): 415-433. <https://doi.org/10.1007/S10844-022-00713-9>

[9] Aidoo-Anderson, A., Polychronakis, Y., Sapountzis, S., Kelly, S. (2025). Investigating demand forecasting practices and challenges in Ghana's Manufacturing Pharmaceutical (MPharma) small and medium enterprises (SMEs): Insights and recommendations. *International Journal of Production Research*, 63(21): 7899-7920. <https://doi.org/10.1080/00207543.2025.2508335>

[10] Syberg, M., Polley, L., Deuse, J. (2025). Evaluating sales forecasting methods in make-to-order environments: A cross-industry benchmark study. *Computer Sciences & Mathematics Forum*, 11(1): 1. <https://doi.org/10.3390/CMSF2025011001>

[11] Martinović, M., Dokic, K., Pudić, D. (2025). Comparative analysis of machine learning models for predicting innovation outcomes: An applied AI approach. *Applied Sciences*, 15(7): 3636. <https://doi.org/10.3390/APP15073636>

[12] Žunić, E., Korjenic, K., Hodzic, K., Donko, D. (2020). Application of Facebook's Prophet algorithm for successful sales forecasting based on real-world data. *International Journal of Computer Science & Information Technology (IJCSIT)*, 2(2): 23-36.

- <https://doi.org/10.5121/ijcsit.2020.12203>
- [13] Petropoulos, F., Makridakis, S., Assimakopoulos, V., Nikolopoulos, K. (2014). 'Horses for courses' in demand forecasting. *European Journal of Operational Research*, 237(1): 152-163. <https://doi.org/10.1016/j.ejor.2014.02.036>
- [14] Guo, L., Fang, W., Zhao, Q., Wang, X. (2021). The hybrid PROPHET-SVR approach for forecasting product time series demand with seasonality. *Computers & Industrial Engineering*, 161: 107598. <https://doi.org/10.1016/J.CIE.2021.107598>
- [15] Giannopoulos, P.G., Dasaklis, T.K., Tsantilis, I., Patsakis, C. (2025). Machine learning algorithms in intermittent demand forecasting: A review. *International Journal of Production Research*, 1-43. <https://doi.org/10.1080/00207543.2025.2578701>
- [16] Zhou, X., Ye, J., Zhao, S., et al. (2026). Noise-robust and sector-aware representation learning for natural gas demand forecasting. *Expert Systems with Applications*, 299: 129964. <https://doi.org/10.1016/J.ESWA.2025.129964>
- [17] Sánchez-Úbeda, E.F., Berzosa, A. (2007). Modeling and forecasting industrial end-use natural gas consumption. *Energy Economics*, 29(4): 710-742. <https://doi.org/10.1016/J.ENECO.2007.01.015>
- [18] Qin, D., Liu, G., Li, Z., Guan, W., Zhao, S., Wang, Y. (2023). Federated deep contrastive learning for mid-term natural gas demand forecasting. *Applied Energy*, 347: 121503. <https://doi.org/10.1016/J.APENERGY.2023.121503>
- [19] Pu, Y., Zhu, C., Yao, G., Yang, K., Yang, Q., Chen, C.L.P. (2026). Interval forecast of natural gas daily consumption based on spatial-temporal Bayesian model. *Advanced Engineering Informatics*, 72: 104426. <https://doi.org/10.1016/J.AEI.2026.104426>
- [20] Syntetos, A.A., Babai, Z., Boylan, J.E., Kolassa, S., Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, 252(1): 1-26. <https://doi.org/10.1016/j.ejor.2015.11.010>
- [21] Hyndman, R.J., Athanasopoulos, G. (2021). 9.9 Seasonal ARIMA models. In *Forecasting: Principles and Practice* (3rd ed.). <https://otexts.com/fpp3/seasonal-arima.html>.