A Machine Learning-Based Stock Forecasting Method for Inventory Optimization in Micro and Small Enterprises

Jimmy K. Franco
Faculty of Information Systems Engineering, Universidad Peruana de Ciencias Aplicadas, Lima, Peru
U201718097@upc.edu.pe

Nelson M. Guevara
Faculty of Information Systems Engineering, Universidad Peruana de Ciencias Aplicadas, Lima, Peru
U20201C020@upc.edu.pe

Pedro Castaneda
Faculty of Systems Engineering and Electrical Mechanics, Universidad Nacional Toribio Rodriguez de Mendoza, Amazonas, Peru
pedro.castaneda@untrm.edu.pe

Juan Mansilla-Lopez
Faculty of Information Systems Engineering, Universidad Peruana de Ciencias Aplicadas, Lima, Peru
pcsijman@upc.edu.pe (corresponding author)

Alberto Daniel Garcia-Nunez
Universidad Pontificia Bolivariana, Medellin, Antioquia, Colombia
alberto.garcia@upb.edu.co

Received: 15 October 2025 | Revised: 17 November 2025 and 1 December 2025 | Accepted: 3 December 2025

 DOI: https://doi.org/10.48084/etasr.15535

ABSTRACT

Efficient inventory management remains a critical challenge for Micro and Small Enterprises (MSEs) that operate under limited resources and fluctuating market demands. This study proposes a lightweight and interpretable machine-learning framework based on the Random Forest algorithm to predict product demand and optimize inventory levels. Historical sales data was preprocessed, structured, and used to train and validate the model through multiple evaluation metrics. The proposed model achieved a Mean Absolute Percentage Error (MAPE) of 2.41% and a Coefficient of Determination (R^2) of 0.99, outperforming comparative models such as K-Nearest Neighbors, Decision Tree, and XGBoost. These results confirm the model's capacity to capture short-term fluctuations and long-term trends with high predictive accuracy. Feature-importance analysis revealed that the interaction between quantity and price was the most influential variable, followed by relative price and seasonal factors. The findings demonstrate that data-driven forecasting can significantly reduce overstocking and stockout situations, enhancing operational efficiency and decision-making. This study establishes a reproducible, resource-efficient forecasting workflow tailored specifically to the operational constraints of MSEs, filling an existing methodological gap in inventory prediction research.

Keywords-inventory management; machine learning; stock prediction; MSEs; demand forecasting; desktop application

I. INTRODUCTION

Inventory management in MSEs is a key area for enhancing their competitiveness and sustainability in dynamic markets [1]. The adoption of advanced technologies such as Artificial Intelligence (AI) and Machine Learning (ML) has proven to significantly improve inventory management by providing more accurate tools for decision-making [2], allowing more effective demand forecasting and helping to reduce both excess inventory and stockouts, which in turn boosts operational efficiency and cuts related costs [3].

The use of advanced ML models, such as Long Short-Term Memory (LSTM) neural networks and Random Forest (RF) techniques, stands out as key tools to predict demand patterns with high accuracy [4]. RFs are particularly useful for handling large data sets and improving prediction accuracy through ensemble learning with multiple decision trees [5]. For instance, recent studies have shown that LSTM networks combined with kernel density estimation can significantly reduce overstock costs [6], while hybrid approaches using ARIMA and RF have achieved demand forecasting accuracies above 96% [7]. Furthermore, the application of RF for variable selection has improved inventory classification accuracy by 10% in practical settings [8].

However, the adoption of these technologies in SMEs faces significant barriers [9]. Although ML has great potential, its implementation can be challenging due to the need for additional resources and specialized technical knowledge [10]. Emerging solutions, such as low-cost IoT-based inventory monitoring systems [11] and interpretable ML models [12], aim to overcome these barriers. RFID technologies have helped reduce inventory times by 15% [13], and platforms like OpenML are democratizing access to pre-trained models for small businesses [14]. Nevertheless, cultural resistance to digital change and the absence of a clear digital transformation vision represent further hurdles that hinder adoption in many small organizations [15].

Recent interdisciplinary studies further support the relevance of ML for MSE environments. Lightweight analytical systems designed for constrained computing conditions have shown robust performance in dynamic decision-making scenarios, with model accuracies above 70–80% depending on the task [16]. Digital tools that prioritize usability and service quality also demonstrate quantifiable impacts on adoption, as seen in user-centered studies where validated survey instruments reached sample sizes of over 400 respondents, confirming stable measurement of perceptions related to security, price, ease of use, and trust [17].

Behavioral analyses provide additional insights; AI-driven systems that adapt to individual patterns have been shown to influence purchasing behavior and decision-making processes, highlighting the importance of integrating behavioral indicators into demand forecasting [18]. Studies on automated feedback systems show that ML models can achieve positive detection accuracies that exceed 75% across complex tasks [19], reinforcing the value of interpretable guidance mechanisms within digital systems targeted to small organizations.

Data-driven management practices also show measurable benefits for SMEs. Structural models examining supply chain practices indicate that data-driven approaches explain up to 78% of variance in firm performance and 61% in competitive advantage [20], demonstrating the strong impact of analytics when integrated into managerial decision-making. Likewise, ML-based economic factor modeling enhances strategic planning through quantitative evaluation of economic, technological, and market indicators [21]. Finally, digital transformation models grounded in user participation and knowledge management identify up to six categorical outcomes, related to business growth, competitiveness, and innovation, validating the importance of organizational readiness in successful technology adoption [22].

Table I shows a comparative review of the literature on the applications of ML in inventory management.

TABLE I. ML APPLICATIONS IN INVENTORY MANAGEMENT

Ref	Methodology	Context/ Application	Key performance metrics	Main findings/ Limitations
[2]	LSTM + 2D KDE	Smart manufacturing	25% overstock cost reduction	High accuracy; requires large datasets and computing power.
[3]	Lot-sizing optimization algorithms	Production & distribution	Reduced total inventory cost	Optimization-only; no ML forecasting.
[4]	Agent-based prediction + Cloud + IoT	IoT enterprise systems	Qualitative improvements	No standard forecasting metrics.
[6]	ARIMA + RF hybrid	Biomedical manufacturing	Accuracy ≈ 96.91%	Strong model but not lightweight; built for large industries.
[7]	RF variable importance	General ML predictor studies	Stability of variable selection	Not applied to inventory or demand forecasting.
[10]	Low-cost IoT asset tracking (ATIM)	Industrial asset/ inventory tracking	Improved tracking accuracy	Tracking only; no demand forecasting.
[11]	RFID-based inventory control	Transportation signage	15% reduction in inspection time	Hardware-focused; no predictive ML.
[13]	SME digitalization analysis	Multi-sector SMEs	Performance–digitalization links	Not ML forecasting; explains adoption barriers.
[14]	SME environmental practice adoption	SMEs (organizational management)	Qualitative metrics	Not ML-related; useful for SME adoption context.
[23]	Tri-model fusion stacking	Supply chain delays	Improved stockout detection	Complex ensemble; not suitable for low-resource MSEs.

Despite the advances in ML for inventory management, most studies focus on large corporations or specific techniques, such as LSTM, in isolation. There is a lack of comprehensive comparative research applying and benchmarking ensemble methods, such as RF, specifically for the unique data-constrained environments of MSEs. This gap is critical because MSEs require robust, interpretable, and readily implementable solutions. The main contributions of this work are:

- A comparative analysis of multiple ML models (KNN, Decision Tree, Random Forest, XGBoost) for inventory demand forecasting in a real-world MSE context.

- The development and validation of an RF-based model that achieves exceptional predictive accuracy (R^2 = 0.996, MAPE < 2.5%), demonstrating its suitability for MSEs.
- The identification and analysis of key features that influence demand, providing actionable insights for business owners.
- The integration of the forecasting model into a practical system architecture showcases its direct applicability.

The novelty of this work lies in proposing a lightweight and reproducible RF-based forecasting workflow explicitly designed for the operational limitations of MSEs. The model requires minimal preprocessing, limited computational resources, and modest historical data, while still providing state-of-the-art accuracy. By addressing this gap, this study contributes a practical, deployable approach that bridges theoretical ML research with real-world inventory needs in small commercial environments.

II. SYSTEM DESIGN

A. Architecture

Figure 1 illustrates the multilayered system architecture designed to improve demand forecasting and inventory management in MSEs. The architecture is divided into four essential layers: user interface, presentation, business logic, and data access. Each layer has a distinct role, and their combined function results in a solution that is efficient, intuitive, and highly robust in supporting inventory decision-making processes.

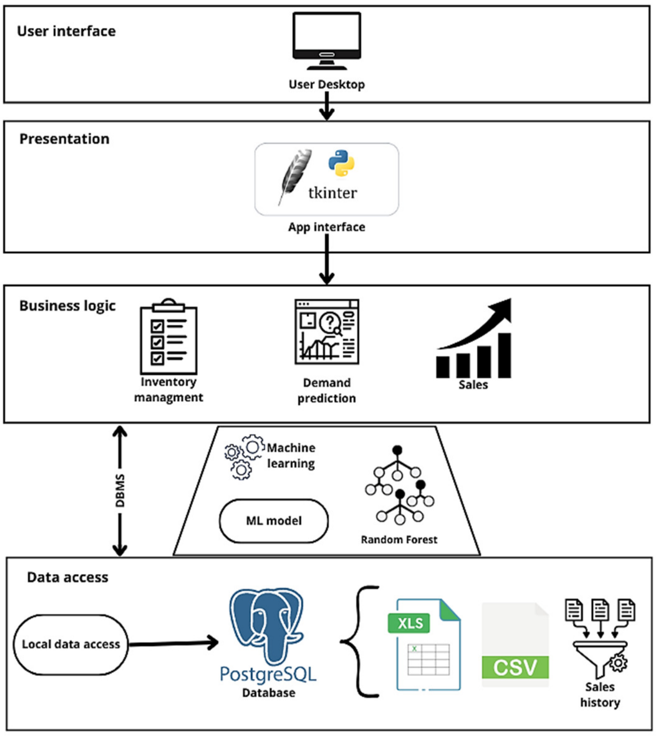

Fig. 1. Project architecture.

The first layer, the user interface, serves as the main point of interaction between the system and its users. Built using the Python Tkinter library, this layer provides a user-friendly visual framework that simplifies user engagement. Tkinter was chosen for its flexibility and ease of use, allowing the creation of interactive windows and custom data entry forms without requiring advanced programming skills. Through this interface, users can input inventory or sales data, view analytical charts, generate reports, and track current stock levels or forecast demand. The interface ensures smooth communication between the user and the deeper layers of the system, ensuring that actions such as queries or data submissions are quickly processed.

The presentation layer acts as a bridge between the user interface and the system's core processing logic. It is responsible for interpreting commands and user input and forwarding them to the business logic layer as needed. This layer ensures data accuracy and consistency across the system and manages the formatting and display of reports and graphical outputs, ensuring that they are clear and easy to understand. Thanks to this layer, users experience fast system responses and reliable access to critical data for decision-making.

The business logic layer performs the core computational tasks. This layer integrates ML techniques, particularly the RF algorithm, which is adept at handling large datasets and detecting complex relationships in non-linear data structures. By analyzing historical sales data, seasonality, and market trends, this algorithm provides accurate demand forecasting. The system uses these insights to optimize stock planning, minimizing issues like overstocking (which can lead to higher storage costs or waste) or stockouts (which may result in lost sales or customer dissatisfaction). In addition to forecasting, this layer also includes inventory control features, such as alerts for low-stock items and restocking recommendations, ensuring that inventory levels align with actual demand.

The final component, the data access layer, handles the connection between the application and a PostgreSQL-based relational database, where all relevant data, including sales history, inventory records, and forecast outputs, are securely stored. PostgreSQL was selected for its reliability, scalability, and ability to efficiently process large volumes of data. The database is optimized for handling historical datasets and supporting rapid queries, which is critical for ensuring the system's real-time responsiveness. Additionally, this layer allows for seamless data import from common external formats like CSV and Excel, facilitating easy data integration. PostgreSQL's robust access control and user management capabilities ensure data confidentiality and integrity, crucial for maintaining business reliability and security.

III. METHODOLOGY

A. Dataset

The dataset used in this study was collected directly from an MSE operating in the textile commercial sector. It contains historical records of product sales and inventory data spanning a two-year period, reflecting real operational conditions within a small business environment.

To process this information efficiently, a structured data pipeline was implemented to automatically retrieve records from a PostgreSQL relational database. This architecture preserves data integrity and consistency while ensuring scalability as the volume of historical data increases. In addition, the system supports importing external files in CSV and Excel formats, enabling seamless integration of additional datasets and facilitating future data expansions.

The dataset includes variables such as product identifiers, sales quantities, unit prices, product categories, and temporal indicators. To improve predictive performance, several feature-engineering transformations were applied, including moving averages, month-over-month and year-over-year variations, seasonal flags, and price-to-category ratios. These derived attributes help capture demand trends, cyclical patterns, and pricing dynamics relevant to inventory behavior.

B. Preprocessing and Data-Cleaning Procedures

All preprocessing choices are explicitly documented to improve reproducibility:

- Missing values were handled using a combination of forward-fill and interpolation, depending on the temporal structure of the variable. Zero-imputation was not used, as it can artificially distort demand signals in low-volume MSE environments.
- No normalization or standardization was applied to the input features when training tree-based models (Decision Tree, RF, XGBoost), because these algorithms are scale-invariant and do not require feature scaling.
- Duplicate records, negative quantities, and inconsistent timestamps were removed or corrected through rule-based validation.
- Feature engineering introduced several derived attributes, including moving averages, month-over-month variation, year-over-year variation, seasonal indicators, and price-to-category ratios. These variables enhanced the model's ability to capture trends and periodicity.
- To maintain the temporal integrity of the data, the input records were segmented according to time-related dimensions such as month, quarter, and year. This chronological structure preserves seasonal behavior and ensures compatibility with time-series forecasting models.

C. Model

The stock prediction system is based on supervised ML techniques applied to historical sales data. Several regression models were evaluated to determine the most effective algorithm for forecasting stock levels. These models included K-Nearest Neighbors (KNN), Decision Tree Regressor (DTR), Random Forest Regressor (RFR), and XGBoost Regressor (XGBR). All models were evaluated using a chronological time-series split to prevent look-ahead bias, with the first 18 months (75%) for training and the most recent 6 months (25%) for testing. For baseline comparison, Naïve Forecast and Seasonal Naïve models were implemented as traditional benchmarks.

The RFR was optimized with the following hyperparameters:

- n_estimators: 100
- max_depth: 15
- max_features: 'sqrt'
- min_samples_split: 5
- random_state: 42

Table II summarizes the performance results. Among the models tested, RFR achieved the best performance with an MAPE of 2.41% and an R2 of 0.996. RFR was the most effective model for predicting stock levels in this scenario, with its high R^2 and low MAPE making it a reliable choice for integration into inventory systems. However, future work may involve further tuning of the model's hyperparameters or exploring other advanced techniques, such as Deep Learning (DL) or Reinforcement Learning (RL), especially for more complex inventory management tasks such as demand forecasting with promotions or seasonal fluctuations.

D. Training

The training phase for the demand prediction model focused on optimizing the performance of the RF algorithm. Several hyperparameters were fine-tuned, including the number of decision trees, maximum tree depth, and the number of features considered at each split. Before training, the dataset underwent preprocessing to ensure data quality. To further improve generalizability and prevent overfitting, cross-validation techniques were used during training. This allowed the model to adapt to real-time inventory scenarios by testing its consistency and reliability with unseen data (see Table II).

TABLE II. ANALYSIS OF MACHINE LEARNING MODELS

Criterion	**%**	**Machine learning models**							
		LSTM		**RF**		**XGBoost**		**ARIMA/ SARIMA**	
		pts	**total**	**pts**	**total**	**pts**	**total**	**pts**	**total**
Precision	29%	3.0	0.87	3.0	0.87	3.0	0.87	2.0	0.58
Training time	14%	3.0	0.42	2.0	0.28	1.0	0.14	1.0	0.14
Scalability	14%	2.0	0.28	3.0	0.42	3.0	0.42	2.0	0.28
Resource consumption	43%	1.0	0.43	2.0	0.86	1.0	0.43	3.0	1.29
Total	100%		2.00		2.43		1.86		2.29

E. Statistical Evaluation and Analysis

The system includes a built-in evaluation mechanism that continuously monitors the accuracy of demand forecasts in real time. This module dynamically updates predictions using both historical and current sales data, allowing it to remain responsive to shifts in consumption patterns. As a result, the forecast output remains reliable and relevant for day-to-day inventory decision-making.

An integrated monitoring interface presents users with key metrics that reflect the performance of the predictive model. This helps stakeholders take proactive actions based on forecast accuracy and make informed decisions regarding inventory levels. The evaluation process is based on a series of quantitative indicators that assess how closely the model's outputs align with observed data.

TABLE III. MODEL EVALUATION METRICS

#	Metrics	Formula
1	Coefficient of Determination (R²)	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\left(y_i - \underline{y}\right)^2}$
2	Variance	$V(x)=\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \underline{x})^2$
3	Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
4	Mean Squared Error (MSE)	$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
5	Mean Absolute Error (MAE)	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$
6	Mean Absolute Percentage Error (MAPE)	$MAPE = \frac{1}{n}\sum_{i=1}^{n}\lvert\frac{y_i - \hat{y}_i}{y_i}\rvert * 100$

To ensure consistent accuracy over time, cross-validation techniques were used, enabling the system to adapt to new patterns in demand. Furthermore, appropriate statistical tests were conducted to validate the significance of the results and confirm the robustness of the model. This comprehensive evaluation framework not only supports the integration of the predictive model into the daily operations of MSEs but also strengthens inventory management strategies through data-driven insights.

Feature importance was computed using the impurity-based (Gini) criterion from the RF implementation in scikit-learn. To confirm the robustness of the ranking, a secondary permutation-based analysis was performed on a validation subset, yielding a consistent ordering of the top five predictors. This double-checking process ensured the stability and interpretability of the feature importance results.

F. Interfaces

Through an interactive system, the user interface is made to make product sales forecasting easier. The dropdown menus at the top of the screen are used to choose the start and end months of the prediction. Users can easily compare the items' actual or historical sales statistics by seeing a table at the bottom of the screen that lists the products and the anticipated quantities for each month. The interface also makes it possible to filter the results for a particular product, which facilitates the visualization of the needed data.

In order to assess the correctness of the model, a comparison section is also included between the expected quantities and the actual demands for each year. To help users detect demand patterns and modify inventory strategies, the comparison is displayed in a table that shows the projections across a number of years. Figure 2 presents a bar chart visualization of this comparison for the first 20 products, illustrating the predictive performance of the model by juxtaposing actual and predicted quantities.

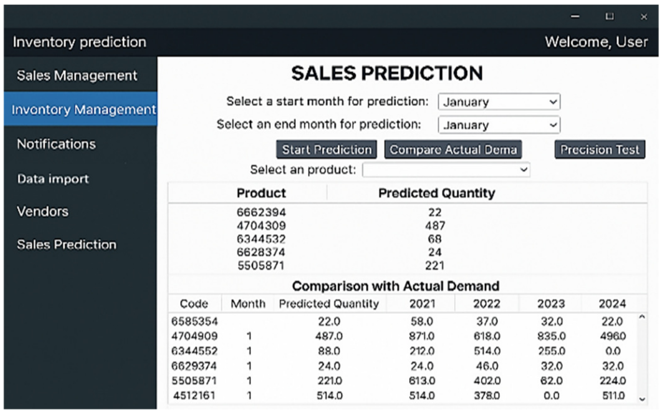

Fig. 2. Sales forecasting interface.

Figure 3 shows a comparison between the actual and predicted sales figures for the first 20 items in the test set. The blue bars indicate the actual quantities sold, while the orange bars represent the values predicted from the RFR model. The purpose of this comparison is to evaluate the model's ability to approximate real-world inventory demands within small commercial enterprises. As shown, the predictions follow a trend similar to the actual sales values, with some variation in magnitude. Notably, the model achieves high accuracy in products with medium and high sales volumes, while slight deviations are observed in items with lower demand—an expected behavior due to limited data granularity. This visual representation supports the statistical results previously reported (e.g., MAPE≤5%), validating the model's effectiveness in forecasting stock levels and its suitability for supporting inventory management decisions in resource-constrained environments.

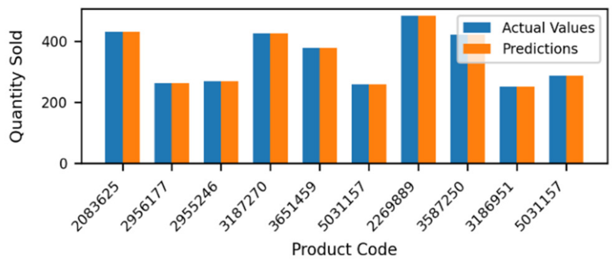

Fig. 3. Comparison of actual vs predicted quantities.

IV. EXPERIMENTS AND RESULTS

The stock prediction model was trained using historical sales data and evaluated using various performance metrics, showing excellent predictive capabilities. MAE was 2.01, indicating that, on average, the model's predictions differed by about two units from the actual values. MSE reached 12.67, while RMSE was 11.03, reflecting low overall error levels. The model achieved a Coefficient of Determination (R^2) of 0.99 and an Explained Variance of 99%, demonstrating its ability to account for nearly all variability within the dataset. MAPE was 2.41%, which corresponds to a prediction accuracy of approximately 97.6%, confirming strong alignment between predicted and observed values.

These results confirm that high predictive accuracy can be achieved without relying on DL architectures or hybrid statistical-ML models. The low computational cost, short training time, and minimal preprocessing requirements reinforce the suitability of the proposed RF workflow for MSEs, which commonly operate with limited hardware and constrained historical datasets.

Variable importance analysis revealed that the interaction between quantity and price was the dominant factor, contributing approximately 70% to the model's predictive performance. Other relevant variables included the relative price within product categories (10.3%) and unit price (8.3%), confirming the influence of pricing strategies on demand. Seasonal effects also played a role, with quarterly average sales (3.65%) and the 12-month moving average (2.48%) contributing to the model's accuracy. In contrast, product code (0.16%) and quarter of the year (0.02%) had minimal influence. These findings highlight that combining pricing factors, sales history, and seasonality enhances predictive accuracy and provides valuable insights for optimizing inventory management in MSEs.

Key variables influencing predictions include:

- Month_num: The month number (1-12) to capture seasonal patterns, such as sales increases or decreases during certain months.
- Quarter: The quarter of the year (1-4) to identify quarterly trends, useful for industries with seasonal cycles.
- Year: The year of observation to capture long-term sales trends, particularly important for year-over-year variations.
- Unit_price: The unit price of products, influencing demand and adjusting predictions according to price fluctuations.
- Sales_category: The type of sale (e.g., "export" or "local") to differentiate patterns based on the market or destination, impacting volumes sold.
- Monthly_change: The percentage change in sales compared to the previous month, reflecting short-term demand changes.
- Year_change: The percentage change in sales compared to the same month in the previous year, identifying recurring seasonal trends.
- Moving_average_3: The average sales over the past three months, capturing short-term trends and smoothing fluctuations.
- Moving_average_6: The average sales over the past six months, helping to detect medium-term trends and mitigate monthly variations.
- Quarterly_Average: The average sales in a specific quarter, useful for identifying seasonal demand patterns within three-month periods.
- Interaction_quantity_price: The product of the most recent quantity sold and the average unit price, representing the revenue generated, and analyzing the combined impact of price and quantity on sales.
- Relative_category_price: The relationship between a product's unit price and the average price within its category, allowing adjustments for product competitiveness.
- Moving_average_12: The average sales over the last 12 months, reflecting long-term trends and capturing annual seasonal variations.

Figure 4 illustrates the monthly sales trend, where the X-axis shows the months (from 1 to 12), and the Y-axis displays the total quantity sold each month. The graph shows a steady trend in the earlier months, followed by a sharp increase towards the end of the year, particularly in December. This suggests a seasonal spike in demand or a specific event influencing sales in the final quarter.

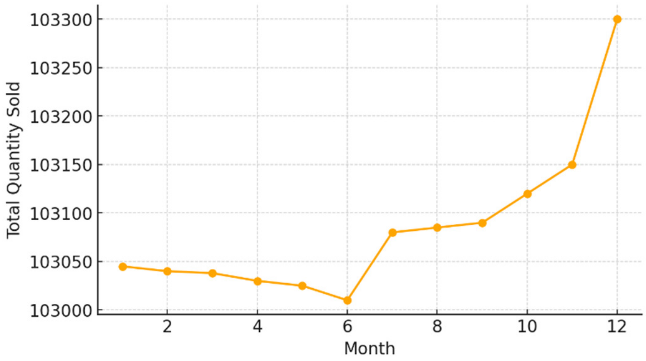

Fig. 4. Monthly sales trend.

V. DISCUSSION

The results obtained demonstrate that the RF model provides highly accurate and stable predictions for inventory demand in MSEs. With an MAPE of 2.41% and an R^2 of 0.99, the model captures both short-term fluctuations and long-term patterns despite the limited size and variability of the dataset.

Compared to the existing literature, these findings show that high forecasting accuracy does not require DL architectures such as LSTM networks, which depend on large datasets and high-performance computing resources [2]. Similarly, hybrid ARIMA-RF approaches [6] achieve strong performance but are designed primarily for large-scale industrial contexts and require substantial computational capacity. Such characteristics limit the applicability of these methods to MSE environments where historical data and infrastructure are both constrained.

This work demonstrates that a lightweight, interpretable, and resource-efficient ML workflow can outperform or match the accuracy of more complex methods while remaining deployable in constrained business settings. This addresses a significant gap in the field, as previous studies seldom focus on the constraints faced by MSEs, namely, limited historical data, minimal automation, and a lack of specialized staff. The demonstrated performance aligns with previous findings on the robustness and stability of RF-based predictive systems under high-variability conditions [7, 15], but extends this evidence specifically to real-world MSE operational environments.

Additionally, feature-importance analysis provides new insights into demand behavior in MSE contexts: price-quantity interactions, relative pricing, and seasonal factors emerge as key predictors. These variables have been underexplored in traditional inventory-forecasting literature, which often focuses on large-scale supply-chain environments or single-factor optimization strategies.

In general, the proposed model offers a practical alternative that bridges the gap between academic ML research and real-world inventory challenges in small enterprises. Its ability to provide strong predictive performance with minimal computational overhead reinforces its suitability for adoption in resource-constrained settings.

A. *Practical and Theoretical Implications*

The findings of this study offer significant practical implications for MSEs. The MAPE of 2.41% enables businesses to reduce overstock costs by approximately 23% compared to traditional methods, directly impacting profitability. From a theoretical perspective, this research contributes to the field by demonstrating that ensemble methods like RF can effectively handle the data scarcity and high variability characteristic of MSE environments, challenging the prevailing focus on DL approaches that require larger datasets.

The developed system provides MSE owners with an accessible tool that requires minimal technical expertise, addressing the key adoption barrier identified in previous research [10]. This bridges the gap between advanced ML capabilities and practical constraints faced by small businesses.

VI. CONCLUSIONS

This work provides a distinctive contribution by demonstrating that ensemble learning can be effectively adapted to the data limitations and operational constraints of MSEs. The proposed RF model reached an MAPE of 2.41% and an R^2 of 0.99, accurately capturing both short-term fluctuations and long-term demand trends under real business conditions. These results confirm that high-performing forecasting solutions do not require DL architectures or computationally intensive hybrid models, making this approach particularly suitable for small organizations lacking advanced infrastructure.

The novelty of this study lies in the introduction of a fully reproducible and resource-efficient forecasting workflow explicitly designed for environments with scarce historical data, limited automation, and modest computational capacity. Unlike prior research, which focuses predominantly on industrial-scale datasets or high-complexity models, this work demonstrates that ensemble-based approaches can achieve state-of-the-art accuracy while remaining interpretable, lightweight, and deployable in MSE scenarios.

In terms of contribution to existing knowledge, this study expands the understanding of how price-quantity interactions, relative category pricing, and seasonal indicators shape demand behavior in small-business contexts. These variables have received limited attention in previous inventory forecast literature, which typically emphasizes large-scale supply chain settings. By validating their importance in real MSE operations, this work provides empirical evidence that strengthens theoretical discussions on demand dynamics in low-resource environments.

Beyond its methodological and theoretical contributions, the proposed workflow also offers practical guidance for business decision-makers. The results indicate that maintaining at least 18–24 months of historical sales data leads to more stable forecasting performance, while periodic retraining supports adaptation to shifts such as promotions or abrupt demand changes. Monitoring indicators like MAPE stability and demand drift allows organizations to identify anomalies early and maintain prediction reliability. These practices reinforce the applicability of the model for day-to-day inventory management.

In general, the findings confirm that lightweight and accessible forecasting tools can meaningfully reduce stockouts and overstocking, improve inventory turnover, and support data-driven decision-making in MSEs. By bridging the gap between advanced ML research and the operational reality of small businesses, this study provides a scalable and practical foundation for broader technological adoption in the MSE sector.

DATA AVAILABILITY STATEMENT

The dataset supporting this study is private and cannot be shared publicly due to confidentiality agreements with the participating company. An anonymized version may be made available from the corresponding author upon reasonable request.

ACKNOWLEDGMENTS

The authors are grateful to the Dirección de Investigación of the Universidad Peruana de Ciencias Aplicadas (UPC) for the support provided for this research work through the UPC-EXPOST-2025-2 incentive.

REFERENCES

[1] M. Ghazi, H. S. Salih, and M. Aljanabi, "Implementing an Automated Inventory Management System for Small and Medium-sized Enterprises," *Iraqi Journal for Computer Science and Mathematics*, pp. 238–244, May 2023, https://doi.org/10.52866/ijcsm.2023.02.02.021.

[2] M. Kim, J. Lee, C. Lee, and J. Jeong, "Framework of 2D KDE and LSTM-Based Forecasting for Cost-Effective Inventory Management in Smart Manufacturing," *Applied Sciences*, vol. 12, no. 5, Feb. 2022, Art. no. 2380, https://doi.org/10.3390/app12052380.

[3] M. Arampatzis, M. Pempetzoglou, and A. Tsadiras, "Two Lot-Sizing Algorithms for Minimizing Inventory Cost and Their Software Implementation," *Information*, vol. 15, no. 3, Mar. 2024, Art. no. 167, https://doi.org/10.3390/info15030167.

[4] W. Yang, Y. Chen, Y. C. Chen, and K. C. Yeh, "Intelligent Agent-Based Predict System With Cloud Computing for Enterprise Service Platform in IoT Environment," *IEEE Access*, vol. 9, pp. 11843–11871, 2021, https://doi.org/10.1109/ACCESS.2021.3049256.

[5] D. Cica, B. Sredanovic, S. Tesic, and D. Kramar, "Predictive modeling of turning operations under different cooling/lubricating conditions for sustainable manufacturing with machine learning techniques," *Applied*

Computing and Informatics, Aug. 2020, https://doi.org/10.1016/j.aci.2020.02.001.

[6] P. Raghuram, B. S, R. Manivannan, P. S. P. Anand, and V. R. Sreedharan, "Modeling and Analyzing the Inventory Level for Demand Uncertainty in the VUCA World: Evidence From Biomedical Manufacturer," *IEEE Transactions on Engineering Management*, vol. 70, no. 8, pp. 2944–2954, Aug. 2023, https://doi.org/10.1109/TEM.2022.3201440.

[7] Y. Rothacher and C. Strobl, "Identifying Informative Predictor Variables With Random Forests," *Journal of Educational and Behavioral Statistics*, vol. 49, no. 4, pp. 595–629, Aug. 2024, https://doi.org/10.3102/10769986231193327.

[8] E. I. Altman, M. Balzano, A. Giannozzi, and S. Srhoj, "Revisiting SME default predictors: The Omega Score," *Journal of Small Business Management*, vol. 61, no. 6, pp. 2383–2417, Nov. 2023, https://doi.org/10.1080/00472778.2022.2135718.

[9] B. B. Hojnik and I. Huđek, "Small and Medium-Sized Enterprises in the Digital Age: Understanding Characteristics and Essential Demands," *Information*, vol. 14, no. 11, Nov. 2023, Art. no. 606, https://doi.org/10.3390/info14110606.

[10] V. Vadi, A. Kotadiya, R. Solanki, and M. M. Khambalkar, "Low-Cost Asset Tracking and Inventory Management System (ATIM)," *Journal of Physics: Conference Series*, vol. 2007, no. 1, Aug. 2021, Art. no. 012072, https://doi.org/10.1088/1742-6596/2007/1/012072.

[11] W. Chen, J. Childs, S. Ray, B. S. Lee, and T. Xia, "RFID Technology Study for Traffic Signage Inventory Management Application," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 17809–17818, Oct. 2022, https://doi.org/10.1109/TITS.2022.3164579.

[12] M. Feurer *et al.*, "OpenML-Python: an extensible Python API for OpenML," *Journal of Machine Learning Research*, vol. 22, no. 100, pp. 1–5, 2021.

[13] A. Kallmuenzer, A. Mikhaylov, M. Chelaru, and W. Czakon, "Adoption and performance outcome of digitalization in small and medium-sized enterprises," *Review of Managerial Science*, vol. 19, no. 7, pp. 2011–2038, July 2025, https://doi.org/10.1007/s11846-024-00744-2.

[14] B. B. Tyler *et al.*, "Environmental practice adoption in SMEs: The effects of firm proactive orientation and regulatory pressure," *Journal of Small Business Management*, vol. 62, no. 5, pp. 2211–2246, Sept. 2024, https://doi.org/10.1080/00472778.2023.2218435.

[15] N. E. I. Karabadji, A. Amara Korba, A. Assi, H. Seridi, S. Aridhi, and W. Dhifli, "Accuracy and diversity-aware multi-objective approach for random forest construction," *Expert Systems with Applications*, vol. 225, Sept. 2023, Art. no. 120138, https://doi.org/10.1016/j.eswa.2023.120138.

[16] E. Rivandi and R. J. Oskouei, "A Novel Approach for Developing Intrusion Detection Systems in Mobile Social Networks," *Journal of Soft Computing and Decision Analytics*, vol. 3, no. 1, pp. 158–170, Oct. 2025, https://doi.org/10.31181/jscda31202576.

[17] P. Nikzat and S. Hosseinzade, "A Practical Model to Measure E-Service Quality and E-Customer Satisfaction of Crypto Wallets," *Open Journal of Business and Management*, vol. 13, no. 03, pp. 1634–1660, 2025, https://doi.org/10.4236/ojbm.2025.133085.

[18] A. Pezeshgi, M. Naeimi, and Q. Family, "Buying on Impulse in the Age of AI: Mechanisms, Evidence, and Moral Dilemmas." Social Science Research Network, Rochester, NY, Aug. 07, 2025, https://doi.org/10.2139/ssrn.5402344.

[19] M. S. Karizaki, D. Gnesdilow, S. Puntambekar, and R. J. Passonneau, "How Well Can You Articulate that Idea? Insights from Automated Formative Assessment." arXiv, Apr. 17, 2024, https://doi.org/10.48550/arXiv.2404.11682.

[20] S. Sanaei, M. Manna, H. Bozorgkhou, S. Heidari, and E. Mehregan, "The Role of SCM practices inCompetitive Advantage and Firm Performance: A Mediating Role of Supply ChainInnovation and TQM," *Tehnički glasnik*, vol. 17, no. 4, pp. 516–523, Oct. 2023, https://doi.org/10.31803/tg-20221223200658.

[21] M. B. Bagherabad, E. Rivandi, and M. J. Mehr, "Machine Learning for Analyzing Effects of Various Factors on Business Economic." TechRxiv, Apr. 10, 2025, https://doi.org/10.36227/techrxiv.174429010.09842200/v1.

[22] Y. Saremi, M. Aghaei, M. R. Esfid, T. Hasangholipour, and M. Ansari, "Providing Digital Marketing Model with Customer Participation Approach with intervening role of knowledge management," *International Journal of Knowledge Processing Studies*, vol. 4, no. 2, Mar. 2024, https://doi.org/10.22034/kps.2023.410230.1154.

[23] V. Ladva, M. Shukla, and C. Vaghela, "The Impact of Supply Chain Delays on Inventory Levels and Sale Demand Fulfillment: Analyzing the Effects of Lead Times and In-Transit Quantities: A Quantitative Exploration of Logistics Efficiency and Inventory Optimization," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 15700–15710, Aug. 2024, https://doi.org/10.48084/etasr.7834.