ISSN 2579-7069

Forecasting Coffee Sales Using Time-Based Features and Machine Learning Models

Yoana Sonia Wijaya[1,*] Ariel Christopher Wawolangi[2]

[1,2]*Department of Information Systems, Faculty of AI and Data Science, Universitas Pelita Harapan, Indoensia*

(Received April 14, 2025; Revised August 18, 2025; Accepted November 20, 2025; Available online January 30, 2026)

Abstract

Sales forecasting is a critical component of operational and strategic decision-making in retail and coffee businesses, where demand exhibits strong temporal variability and product-level heterogeneity. Accurate hourly-level forecasts enable effective inventory management, workforce scheduling, and data-driven promotional strategies. However, existing studies predominantly rely on aggregated sales data and provide limited comparative analyses between traditional statistical models and machine learning approaches using real transaction-level data. This study addresses this gap by conducting an empirical comparison between a traditional ARIMA model and ensemble machine learning models, namely Random Forest and XGBoost, for hourly coffee sales forecasting. The analysis is based on a real-world dataset comprising 3,547 transaction records enriched with temporal and product-related features. Model performance was evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). The results demonstrate that machine learning models significantly outperform the ARIMA baseline, with XGBoost achieving the best performance and explaining approximately 83% of the variance in sales data, while ARIMA shows limited explanatory power due to its inability to capture non-linear and highly volatile demand patterns. Feature importance analysis further reveals that product-specific attributes are the dominant drivers of sales predictions, complemented by seasonal and intra-day temporal effects. These findings provide both scientific and practical contributions by offering empirical evidence on the superiority of machine learning models for granular sales forecasting and supporting data-driven decision-making in coffee retail analytics.

Keywords: Sales Forecasting, Machine Learning, Time Series Analysis, Coffee Retail, Data Analytics

1. Introduction

Sales forecasting plays a critical role in operational and strategic decision-making within the retail and food service industries. Accurate sales predictions enable organizations to optimize inventory management, balance workforce allocation, and improve logistics planning, thereby enhancing operational efficiency and customer satisfaction [1], [2]. Advanced forecasting techniques reduce uncertainty and provide actionable insights that support warehouse capacity management and transportation planning, ultimately strengthening overall supply chain performance [2]. Furthermore, precise sales forecasts inform marketing, financial planning, and supply chain strategies, enabling businesses to respond more effectively to fluctuating consumer demand and dynamic market conditions [3].

In retail contexts such as grocery, e-commerce, and food services, sales forecasting is essential for minimizing costs associated with overstocking and stockouts while ensuring efficient resource allocation [3]. Recent studies highlight that the adoption of machine learning and big data analytics significantly improves forecasting accuracy, providing retailers with a competitive advantage in increasingly data-driven markets [1], [4], sales forecasting has become a fundamental component of modern retail analytics, supporting both short-term operational decisions and long-term strategic planning.

Despite these advances, Small and Medium-Sized Enterprises (SMEs), particularly coffee businesses, often rely on intuition-based forecasting and historical heuristics rather than analytical models. Such traditional approaches are prone to substantial inaccuracies, as they fail to adequately capture complex demand patterns and rapidly changing consumer behaviors [5], [6]. As a result, SMEs frequently face challenges such as excess inventory, stockouts, and missed sales

***Corresponding author: Yoana Sonia Wijaya (01081210001@student.uph.edu)**
DOI: https://doi.org/10.47738/ijiis.v9i1.294

opportunities, which can undermine profitability and operational stability. These limitations are further exacerbated by constrained resources and limited access to advanced data analytics capabilities, placing SMEs at a disadvantage compared to larger organizations that leverage sophisticated forecasting technologies [6].

Advances in time series analysis and machine learning have substantially transformed sales forecasting methodologies. Traditional statistical models, such as Autoregressive Integrated Moving Average (ARIMA) and exponential smoothing, remain widely used due to their interpretability and robustness. However, they are increasingly complemented or replaced by machine learning models capable of capturing non-linear relationships and complex interactions in sales data [7]. Recent studies demonstrate that hybrid approaches integrating statistical and machine learning techniques can further enhance forecasting accuracy by leveraging the strengths of both paradigms [8].

Moreover, deep learning architectures, including Long Short-Term Memory (LSTM) networks and Temporal Fusion Transformers, have shown promise in modeling temporal dependencies and high-dimensional datasets [7], [8]. These models enable more nuanced demand forecasting by incorporating multiple exogenous factors, such as seasonality, promotions, and consumer behavior patterns [9]. Complementary preprocessing techniques, such as Variational Mode Decomposition, have also been introduced to mitigate noise and enhance prediction quality, further improving forecasting performance [10]. Collectively, these developments underscore a paradigm shift toward data-driven and intelligent forecasting frameworks in retail analytics.

Nevertheless, a significant research gap remains in the sales forecasting literature. Most existing studies rely on daily or aggregated sales data, overlooking the richer insights offered by hourly transaction-level data. Granular temporal data can reveal intra-day consumption patterns and short-term demand fluctuations that are critical for operational decision-making, particularly in coffee retail settings. While Hapsani [11] explored hourly sales forecasting using Seasonal ARIMA and linear regression, comprehensive investigations employing advanced machine learning models with such granular data remain limited. Furthermore, rigorous comparative analyses between traditional statistical time series models and machine learning regressors are scarce. Several studies focus on related but distinct aspects of sales analytics without directly evaluating the relative strengths and weaknesses of these approaches in forecasting contexts [12], [13].

To address these gaps, this study aims to investigate sales forecasting in small and medium-sized coffee businesses using hourly transaction-level data and to conduct a systematic comparison between traditional statistical time series models and modern machine learning regressors. Specifically, the objectives of this research are threefold: (1) to examine the impact of hourly data granularity on forecasting accuracy, (2) to compare the predictive performance of statistical models (e.g., ARIMA and SARIMA) with machine learning approaches (e.g., Random Forest and LSTM), and (3) to assess the influence of external factors, such as promotional activities and weather conditions, on forecasting accuracy across different modeling paradigms.

The contributions of this study are fourfold. First, it proposes an enhanced forecasting framework that leverages hourly transaction-level data to capture fine-grained temporal demand patterns. Second, it provides a comprehensive comparative analysis of statistical and machine learning-based forecasting models, offering empirical insights into their relative effectiveness in retail contexts. Third, it identifies key temporal and external factors influencing coffee sales, contributing to a deeper understanding of demand dynamics. Finally, the findings offer practical implications for SMEs in the coffee industry, supporting data-driven decision-making for inventory management, operational planning, and customer satisfaction enhancement.

2. Literature Review

2.1. Sales Forecasting in Retail and Food Service Industries

Sales forecasting has been widely recognized as a fundamental component of operational and strategic decision-making in the retail and food service industries. Accurate sales predictions enable businesses to optimize inventory management, reduce the risks of overstocking and stockouts, and enhance customer satisfaction through more responsive marketing and service strategies. Prior studies highlight that the increasing availability of digital transaction data has accelerated the adoption of data-driven forecasting approaches, transforming traditional supply chain and

 ISSN 2579-7069

inventory management practices [14]. In particular, analytics-driven forecasting has been shown to improve demand responsiveness by aligning product offerings with dynamic consumer behavior, thereby strengthening overall operational efficiency in retail environments.

Recent research has increasingly focused on the application of machine learning techniques to address the limitations of conventional statistical forecasting models. In the food service sector, deep learning approaches such as Recurrent Neural Networks (RNNs) have demonstrated strong predictive capabilities by effectively modeling temporal dependencies in sales data. Manoj [15] reports that RNN-based models can leverage large-scale and high-dimensional datasets to generate accurate restaurant sales forecasts, underscoring their suitability for real-world business applications. Similarly, studies in retail e-commerce emphasize the role of predictive analytics in maintaining inventory effectiveness and improving customer satisfaction amid rapidly evolving market conditions [16]. These findings collectively suggest that machine learning-based forecasting models offer significant advantages in capturing complex and non-linear demand patterns.

Despite these advances, the literature reveals a notable gap in comprehensive comparative analyses between traditional time series models, such as ARIMA, and modern machine learning approaches. While prior studies often demonstrate the effectiveness of individual forecasting techniques, systematic evaluations that contrast their relative strengths and limitations remain limited. Addressing this gap is essential for guiding practitioners and researchers in selecting appropriate forecasting methods tailored to specific operational contexts. Consequently, further research is needed to provide empirical evidence on the comparative performance of statistical and machine learning-based models in retail and food service sales forecasting.

2.2. Statistical Time Series Models: ARIMA and Related Approaches

Statistical time series models, particularly the ARIMA and its seasonal extension SARIMA, have long been established as foundational approaches for forecasting across various domains, including retail and food service industries. These models rely on historical observations to identify temporal structures such as trends and seasonality, enabling the generation of future demand estimates. ARIMA integrates three key components autoregression, differencing, and moving averages allowing it to model both short-term dependencies and longer-term trends in time series data [17], [18]. Due to its interpretability and theoretical robustness, ARIMA is frequently regarded as a benchmark model for traditional forecasting tasks.

In retail and food service contexts, ARIMA-based models have been extensively applied to sales forecasting and inventory management. Seasonal variants, such as SARIMA, are particularly effective in capturing cyclical demand patterns commonly observed in consumer purchasing behavior. Zhao [18] demonstrates that SARIMA models can accommodate seasonal fluctuations while maintaining reliable forecasting performance, making them well suited for industries with recurring consumption cycles. These characteristics have contributed to the widespread adoption of ARIMA models for operational planning, demand estimation, and resource allocation in retail environments.

Despite their strengths, ARIMA models exhibit notable limitations when applied to complex and highly dynamic datasets. Their reliance on linear assumptions constrains their ability to capture non-linear relationships and abrupt demand shifts, which are increasingly prevalent in modern retail systems. Additionally, the requirement for stationarity often necessitates extensive preprocessing, increasing modeling complexity and reducing adaptability. To address these challenges, recent studies have explored hybrid forecasting frameworks that combine ARIMA with machine learning techniques, such as neural networks, to enhance predictive accuracy and flexibility [19]. These hybrid approaches highlight the continued relevance of ARIMA within broader analytical paradigms while underscoring the need for comparative evaluations against modern machine learning models.

2.3. Machine Learning Methods for Sales Prediction

Machine learning methods have increasingly become integral to sales forecasting, offering substantial improvements over traditional statistical approaches by effectively modeling non-linear relationships and accommodating high-dimensional data. Algorithms such as gradient boosting methods including Extreme Gradient Boosting (XGBoost), Gradient-Boosted Regression Trees (GBRT), and Light Gradient Boosting Machine (LightGBM) have demonstrated strong predictive performance across diverse retail contexts. Chen et al. [20] highlight the effectiveness of GBRT and

LightGBM in forecasting demand for products with short shelf lives, underscoring the suitability of ensemble-based learning techniques for complex retail environments.

Beyond tree-based models, deep learning architectures have shown considerable promise in capturing temporal dependencies inherent in sales data. Recurrent neural networks, particularly LSTM models, as well as more recent architectures such as Temporal Fusion Transformers, have demonstrated superior performance in multivariate and long-horizon forecasting scenarios. Theodoridis and Tsadiras [7] emphasize that these deep learning models outperform traditional time series techniques by learning intricate temporal patterns and interactions among multiple explanatory variables. Furthermore, hybrid approaches that integrate statistical models, such as ARIMA, with machine learning techniques have emerged as effective solutions for improving forecasting accuracy. Studies by Kontopoulou et al. [21] indicate that such hybrid frameworks can significantly reduce forecast errors by combining the interpretability of statistical models with the flexibility of machine learning.

Empirical applications of machine learning-based forecasting further reinforce their practical relevance. In fashion retail, Kızgın et al. [22] demonstrate that machine learning models exhibit greater resilience than traditional methods during periods of market disruption, such as the COVID-19 pandemic. Additionally, Punia [23] highlights the importance of incorporating external factors including promotional activities and economic indicators into machine learning models to enhance demand prediction accuracy. The effectiveness of these approaches is commonly evaluated using performance metrics such as Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), which provide robust measures of predictive accuracy. Sun [24] shows that machine learning models consistently outperform classical statistical methods, particularly in datasets that deviate from strict linear or stationary assumptions. Collectively, these findings underscore the growing dominance of machine learning methods in modern sales forecasting and their capacity to support data-driven decision-making in complex retail and food service contexts.

2.4. Summary of Limitations in Existing Studies and Research Positioning

Despite substantial progress in sales forecasting research through the adoption of statistical and machine learning models, several limitations remain evident in the existing literature. One recurring issue concerns the limited utilization of diverse and comprehensive datasets. Although studies such as Tang [25] demonstrate the value of integrating predictive modeling with customer segmentation techniques, many forecasting frameworks still rely on restricted feature sets that inadequately incorporate demographic attributes or external economic indicators. This constraint hampers the ability of forecasting models to capture personalized consumer behavior and broader market dynamics, thereby limiting their predictive effectiveness across varying retail contexts.

Another notable gap lies in the insufficient emphasis on category-specific sales forecasting and systematic model comparison. While demand forecasting has been widely explored in retail research, many studies overlook variations in forecasting performance across different product categories, despite evidence that demand patterns can differ substantially between categories [26]. Furthermore, although some research evaluates the accuracy of individual forecasting approaches, comprehensive comparative analyses between traditional statistical models such as ARIMA and modern machine learning techniques remain relatively scarce [8]. The absence of rigorous, side-by-side evaluations across consistent datasets restricts practitioners' ability to make informed decisions regarding model selection for specific operational environments.

In addition to methodological limitations, the applicability of existing studies to real-world business settings remains a concern. A number of forecasting studies emphasize theoretical performance improvements without sufficiently addressing practical challenges, such as adaptability to market volatility, rapidly shifting consumer preferences, and operational constraints faced by small and medium-sized enterprises. Positioned within this context, the present study seeks to address these gaps by leveraging granular hourly transaction-level data and conducting a systematic comparative analysis between traditional statistical time series models and advanced machine learning approaches. By focusing on real transaction data and category-specific insights, this research aims to enhance forecasting accuracy while ensuring practical relevance for decision-making in the retail and food service industries.

 ISSN 2579-7069

3. Methodology

3.1. Dataset Description

This study utilizes a real-world coffee sales transaction dataset obtained from a retail coffee shop, representing daily operational activities within the food service sector. The dataset captures point-of-sale transaction records and is designed to reflect actual consumer purchasing behavior in a small-to-medium-sized coffee business environment. All records were anonymized to ensure data privacy, and the dataset contains no personally identifiable customer information.

The dataset consists of 3,547 individual transaction records, collected over a continuous observation period spanning multiple months. Each transaction represents a completed sale and includes detailed temporal information, enabling fine-grained analysis of intra-day and inter-day sales patterns. The dataset is chronologically ordered and contains no missing values, making it suitable for time series modeling and machine learning-based forecasting without the need for extensive data imputation.

The variables in the dataset can be categorized into temporal variables and transactional variables. Temporal variables include the transaction timestamp, hour of day, day of the week, and month, which facilitate the analysis of sales dynamics across different time scales. Transactional variables capture key sales attributes, such as the transaction value (sales amount), product type (coffee category), and payment method. These features enable the modeling of both time-dependent demand patterns and contextual sales characteristics. The combination of granular temporal attributes and transactional information provides a robust foundation for evaluating and comparing statistical time series models and machine learning approaches in sales forecasting tasks.

3.2. Data Preprocessing

Data preprocessing was conducted to ensure data quality, consistency, and suitability for both statistical time series analysis and machine learning models. The initial stage involved data cleaning and validation, including verification of data completeness, detection of duplicate records, and examination of outliers in the transaction value variable. The dataset contained no missing values, and all transaction records were valid, allowing subsequent analysis to proceed without the need for imputation or record removal.

To enable time-aware modeling, a unified timestamp was constructed by combining the date and time attributes of each transaction. The resulting timestamp variable was converted into a standardized datetime format and used to chronologically order all records. This ordering was critical to prevent data leakage during model training and evaluation, particularly for time series forecasting tasks. Temporal consistency was further ensured by validating the continuity of the observation period and aggregating transactions when required for specific modeling approaches, such as hourly-based ARIMA analysis.

Handling of categorical and numerical variables was performed in accordance with the requirements of the applied models. Numerical features, including transaction value and hour of day, were retained in their original scale to preserve interpretability. Categorical variables, such as product type, payment method, day of the week, and time-of-day category, were encoded using one-hot encoding to facilitate their integration into machine learning algorithms. This preprocessing strategy ensured that both temporal and transactional features were appropriately represented, enabling a fair and consistent comparison between traditional statistical models and machine learning-based forecasting approaches.

3.3. Feature Engineering

Feature engineering was performed to extract meaningful representations from the raw transactional data and to enhance the predictive capability of the forecasting models. Given the temporal nature of sales data, particular emphasis was placed on constructing time-based features that capture both intra-day and inter-day demand patterns. These features enable the models to learn recurring consumption behaviors and temporal fluctuations commonly observed in retail and food service environments.

Time-based features were derived from the transaction timestamp and include the hour of day, day of the week, and month of transaction. In addition, a categorical time-of-day variable was introduced to represent broader consumption

 ISSN 2579-7069

periods, such as morning, afternoon, and evening. This abstraction facilitates the identification of general demand trends while reducing noise associated with fine-grained hourly variations. Collectively, these temporal features allow the models to capture cyclical patterns, seasonal effects, and peak demand periods at multiple temporal resolutions.

Transaction-related attributes were incorporated to provide contextual information about each sale. These attributes include the transaction value, product category (coffee type), and payment method, which reflect customer purchasing behavior and operational characteristics. Categorical transaction attributes were encoded using one-hot encoding to ensure compatibility with machine learning algorithms, while numerical attributes were preserved in their original scale to maintain interpretability.

The target variable for the forecasting task was defined as the sales value (monetary amount) per transaction, representing the primary indicator of business performance. For time series-based models, such as ARIMA, the target variable was aggregated at an hourly level to form a univariate time series, whereas for machine learning models, the transaction-level target was retained to leverage the full granularity of the dataset. This dual representation enabled a consistent and fair comparison between statistical time series methods and machine learning-based forecasting approaches.

3.4. Forecasting Models

This study employs both statistical and machine learning-based forecasting models to evaluate their effectiveness in predicting coffee sales. The selected models represent widely adopted approaches in sales forecasting literature and enable a balanced comparison between traditional time series techniques and modern data-driven methods.

As a statistical baseline, the ARIMA model was implemented to establish a reference level of forecasting performance. ARIMA models temporal dependencies by combining autoregressive terms, differencing operations, and moving average components, making them suitable for univariate time series forecasting. In this study, the ARIMA model was applied to hourly aggregated sales data to capture underlying temporal patterns while ensuring stationarity through appropriate differencing. Model order selection was conducted using information criteria, such as the Akaike Information Criterion (AIC), to identify the optimal parameter configuration. The ARIMA baseline provides a transparent and interpretable benchmark against which the performance of machine learning models can be assessed.

For machine learning-based forecasting, two ensemble learning models were adopted: the Random Forest Regressor and XGBoost. Random Forest is an ensemble method that constructs multiple decision trees using bootstrap sampling and feature randomness, thereby reducing variance and improving generalization performance. Its robustness to noise and ability to model non-linear relationships make it well suited for transactional sales data with heterogeneous patterns. XGBoost, a gradient boosting-based algorithm, was selected for its strong predictive capability and computational efficiency. By sequentially minimizing prediction errors through gradient-based optimization, XGBoost effectively captures complex interactions among features. In scenarios where XGBoost implementation is constrained, Gradient Boosting Regression serves as a viable alternative, offering similar advantages in modeling non-linear relationships.

The combination of ARIMA and ensemble-based machine learning models enables a comprehensive evaluation of forecasting performance across different modeling paradigms. This comparative framework facilitates an empirical assessment of the strengths and limitations of each approach in handling granular transaction-level data, thereby providing insights into their suitability for sales forecasting in retail and food service contexts.

3.5. Experimental Setup

The experimental setup was designed to ensure a fair, reproducible, and time-aware evaluation of all forecasting models. Given the temporal nature of sales data, a time-based train–test split strategy was employed instead of random sampling to prevent information leakage from future observations. The dataset was chronologically ordered using the constructed timestamp, after which approximately 80% of the earliest observations were allocated to the training set, while the remaining 20% were reserved for testing. This approach simulates real-world forecasting conditions, where models are trained on historical data and evaluated on unseen future transactions.

 ISSN 2579-7069

Model training procedures were tailored to the specific requirements of each forecasting approach. For the ARIMA model, training was conducted on the hourly aggregated sales series derived from the transaction-level data. Model parameters were selected using an automated search over plausible combinations of autoregressive, differencing, and moving average terms, guided by information criteria such as the Akaike Information Criterion (AIC). This process ensured an optimal balance between model fit and complexity while maintaining interpretability.

For machine learning models, training was performed on transaction-level data using the engineered temporal and transactional features. Hyperparameter configurations for the Random Forest Regressor and XGBoost were selected based on commonly adopted best practices in the literature. Key hyperparameters, including the number of trees, tree depth, and learning rate, were tuned through iterative experimentation on the training set to improve predictive performance while avoiding overfitting. All models were trained under identical data partitioning and preprocessing conditions, enabling a consistent and unbiased comparison of forecasting performance across statistical and machine learning paradigms.

3.6. Evaluation Metrics

The forecasting performance of all models was evaluated using three widely adopted regression metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2). These metrics were selected to provide a comprehensive assessment of prediction accuracy, error magnitude, and explanatory power, ensuring a robust comparison between statistical and machine learning-based forecasting approaches.

RMSE measures the square root of the average squared differences between predicted and actual values. RMSE penalizes larger errors more heavily, making it particularly sensitive to substantial prediction deviations. This characteristic is valuable in sales forecasting contexts where large errors can lead to significant operational inefficiencies, such as inventory shortages or excess stock.

MAE represents the average absolute difference between predicted and observed values. Unlike RMSE, MAE treats all errors uniformly, providing an intuitive interpretation of the typical prediction error in the same unit as the target variable. MAE is therefore well suited for evaluating overall forecasting accuracy without disproportionately emphasizing extreme errors.

R^2 quantifies the proportion of variance in the observed sales values that is explained by the forecasting model. R^2 provides insight into the model's explanatory capability and goodness of fit, with higher values indicating stronger predictive performance. Together, RMSE, MAE, and R^2 offer complementary perspectives on model effectiveness, enabling a balanced and reliable evaluation of forecasting performance across different modeling paradigms.

4. Results and Discussion

4.1. Model Performance Comparison

The quantitative performance comparison of the forecasting models is summarized in Table 1 The results demonstrate a substantial performance gap between the traditional statistical model and machine learning-based approaches.

Table 1. Model Performance Comparison on Test Set

Model	RMSE	MAE	R^2
XGBoost	**1.97**	1.33	**0.83**
Random Forest	1.97	**1.33**	0.83
ARIMA (hourly aggregated)	30.37	18.54	0.08

As shown in Table 1, both machine learning models achieve significantly lower RMSE and MAE values compared to the ARIMA baseline, indicating superior predictive accuracy. The XGBoost model yields the best overall performance with the highest R^2 value (0.83), suggesting that it explains approximately 83% of the variance in sales data. In contrast, the ARIMA model exhibits very limited explanatory power ($R^2 = 0.08$), confirming its inadequacy for modeling highly variable transaction-level sales data.

4.2. Prediction Accuracy Analysis

The prediction accuracy of the forecasting models was further examined through visual comparisons between actual and predicted sales values on the test set. These visualizations provide an intuitive assessment of each model's ability to capture sales dynamics beyond the quantitative evaluation metrics.

Figure 1 presents the actual versus predicted sales generated by the Random Forest model. The visualization shows that the predicted values closely follow the actual sales across most observations, indicating a strong alignment between model output and real transaction values. While minor deviations are observable particularly at lower sales points the Random Forest model successfully captures the general magnitude and variability of sales. This suggests that the model is effective in learning non-linear patterns from transaction-level and temporal features, although slight underestimation occurs during abrupt sales drops.

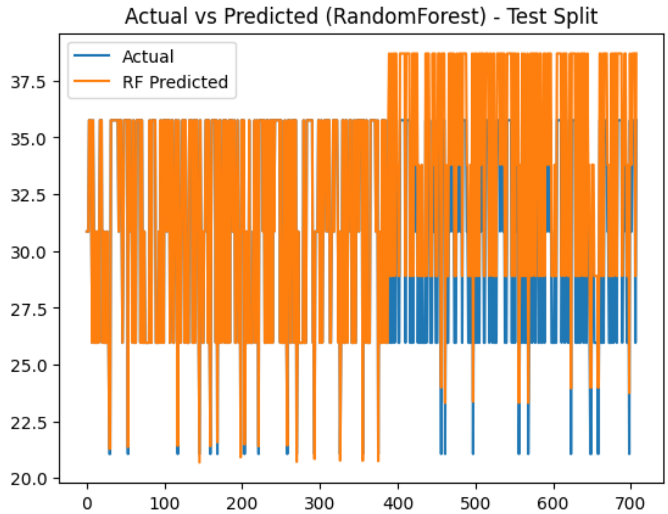

Figure 1. Comparison of Actual and Predicted Coffee Sales Using the Random Forest Model on the Test Set

Similarly, Figure 2 depicts the actual versus predicted sales produced by the XGBoost (or Gradient Boosting) model. Compared to Random Forest, the XGBoost predictions exhibit smoother transitions and reduced volatility, especially during periods of higher sales values. The closer overlap between actual and predicted lines indicates superior generalization performance, which is consistent with the higher R^2 and lower RMSE reported earlier. This behavior confirms the ability of gradient boosting models to better handle complex feature interactions and non-linear demand fluctuations.

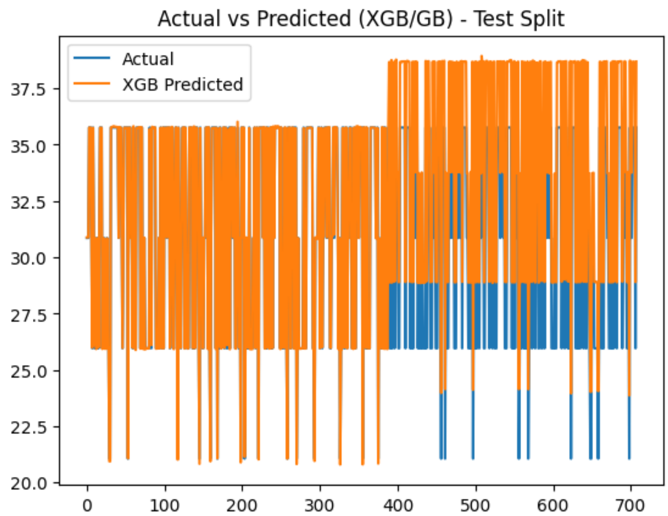

Figure 2. Comparison of Actual and Predicted Coffee Sales Using the XGBoost Model on the Test Set

In contrast, the limitations of the statistical baseline are clearly visible in Figure 3, which illustrates the ARIMA forecast on hourly aggregated sales. The ARIMA predictions remain relatively stable and oscillatory, failing to capture sharp peaks and extreme variations present in the actual sales data. This pattern reflects ARIMA's reliance on linear assumptions and its inability to respond effectively to sudden changes in consumer demand. Consequently, the error magnitude increases substantially during high-demand periods, explaining the model's low explanatory power and high error metrics.

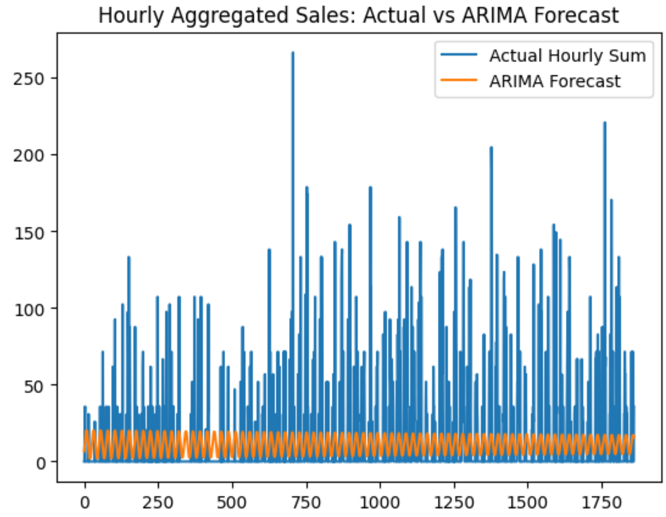

Figure 3. Actual and Forecasted Hourly Aggregated Coffee Sales Using the ARIMA Model

Overall, the error behavior across different time periods indicates that machine learning models maintain more stable and reliable prediction accuracy under varying demand conditions. In particular, the gradient boosting-based approach demonstrates robustness against both low and high sales fluctuations, making it more suitable for real-world deployment in transaction-level sales forecasting.

4.3. Feature Importance Analysis

Feature importance analysis was conducted to identify the most influential factors driving sales predictions and to enhance the interpretability of the machine learning model. As illustrated in Figure 4, the Random Forest model highlights a clear dominance of product-related features, particularly specific coffee types, over purely temporal attributes. Among all features, coffee_name_Cortado, coffee_name_Espresso, and coffee_name_Americano exhibit the highest importance scores, indicating that product preference is a primary determinant of transaction value in the dataset.

Regarding temporal features, monthly indicators such as August, September, and July show moderate importance, suggesting the presence of seasonal demand patterns. The hour of day feature also contributes to the prediction process, although its relative importance is lower compared to product categories and month-based features. This finding implies that while intra-day consumption patterns influence sales, their impact is secondary to product choice and seasonal effects when predicting transaction-level sales values.

The observed feature importance distribution provides insight into the behavior of the Random Forest model. The strong emphasis on product-related attributes demonstrates the model's ability to effectively leverage high-dimensional categorical information through ensemble learning. At the same time, the inclusion of temporal features allows the model to capture broader seasonal and time-based variations without over-relying on them. This balanced utilization of transactional and temporal information explains the model's strong predictive performance and robustness across different time periods. Overall, the results shown in Figure 4 confirm that integrating detailed product information with temporal context is essential for accurate sales forecasting in retail coffee environments.

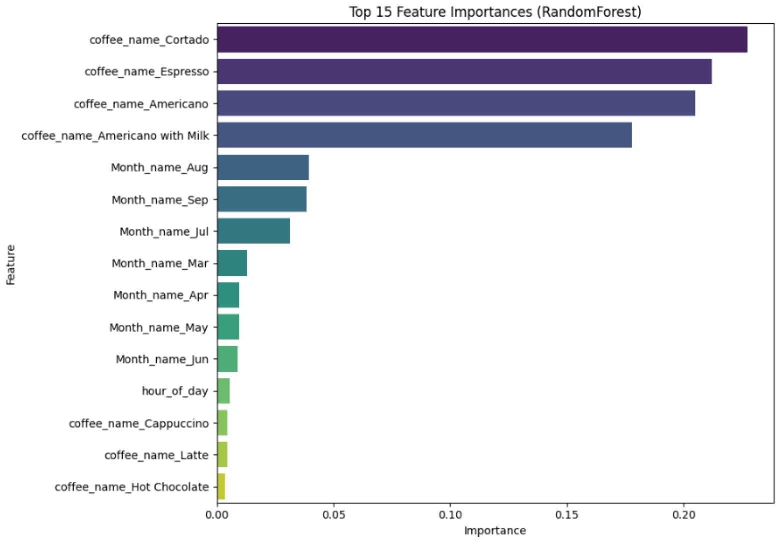

Figure 4. Feature Importance Scores Derived from the Random Forest Model

4.4 Managerial and Business Implications

The findings of this study offer several practical implications for retail coffee businesses, particularly small and medium-sized enterprises. First, the superior accuracy of machine learning models suggests that adopting data-driven forecasting tools can significantly improve inventory management. By accurately predicting sales demand at granular levels, businesses can reduce waste, avoid stockouts, and optimize procurement planning.

Second, insights derived from temporal features can inform staff scheduling and operational planning. Understanding peak sales periods and high-demand products allows managers to allocate workforce resources more efficiently, improving service quality during busy hours while minimizing labor costs during off-peak periods.

Finally, the identification of high-impact product features enables data-driven promotion strategies. Businesses can design targeted promotions for popular coffee types or introduce dynamic pricing strategies during peak demand periods. Seasonal patterns identified through monthly features further support the planning of time-specific marketing campaigns, enhancing customer engagement and revenue potential.

Overall, the results demonstrate that machine learning-based forecasting provides not only superior predictive performance but also actionable insights that support strategic and operational decision-making in retail and food service contexts.

5. Discussion

The results of this study provide strong empirical evidence supporting the growing consensus in the literature that machine learning-based approaches offer substantial advantages over traditional statistical time series models for sales forecasting in retail and food service contexts. Consistent with prior studies emphasizing the importance of data-driven forecasting for inventory optimization and operational efficiency [14], the findings demonstrate that ensemble machine learning models particularly XGBoost and Random Forest achieve significantly higher predictive accuracy than the ARIMA baseline when applied to transaction-level coffee sales data.

The superior performance of machine learning models can be attributed to their ability to capture complex, non-linear relationships inherent in retail sales data. This observation aligns with earlier research reporting the effectiveness of gradient boosting and ensemble-based models in handling heterogeneous and high-dimensional datasets [20]. In contrast, the ARIMA model exhibited limited explanatory power, which corroborates the limitations highlighted in previous studies regarding its reliance on linear assumptions and stationarity requirements [18]. Although ARIMA has long been regarded as a benchmark for traditional forecasting [17], the present results confirm that its applicability diminishes when confronted with highly volatile and granular transaction-level data.

Furthermore, the visual and quantitative analyses reveal that gradient boosting-based models demonstrate greater stability across different demand regimes, including both low and high sales periods. This robustness supports findings from recent literature indicating that machine learning models are more resilient under fluctuating market conditions and disruptions. Unlike ARIMA, which smooths extreme variations and fails to adapt to sudden demand shifts, machine learning models dynamically adjust predictions by leveraging multiple temporal and transactional features. This characteristic is particularly relevant for food service environments, where consumption patterns are influenced by time-of-day effects, seasonality, and product-specific preferences.

The feature importance analysis further enriches the discussion by shedding light on the underlying drivers of sales predictions. The dominance of product-related features, such as specific coffee types, highlights the importance of category-level information in transaction-level forecasting. This finding addresses a key limitation identified in prior studies, where category-specific forecasting has often been overlooked despite its practical relevance [27]. At the same time, the moderate contribution of temporal features, including month and hour-of-day indicators, confirms the presence of seasonal and intra-day demand patterns, as suggested by earlier research on retail consumption behavior [14], [18]. The balanced utilization of product and temporal features illustrates the strength of ensemble learning methods in integrating diverse sources of information, a capability absent in univariate statistical models.

From a broader methodological perspective, this study contributes to the literature by providing a systematic and fair comparison between traditional statistical time series models and modern machine learning approaches using a consistent dataset and evaluation framework. This directly addresses the research gap identified by Martins and Galegale [8], who noted the scarcity of side-by-side evaluations across comparable experimental settings. By leveraging granular hourly transaction-level data, the present work also responds to calls for more realistic and operationally relevant forecasting studies, particularly for small and medium-sized enterprises [25].

Overall, the discussion reinforces the position that machine learning-based forecasting models are not merely incremental improvements over traditional approaches but represent a fundamental shift in how sales demand can be modeled and interpreted in modern retail and food service industries. While statistical models such as ARIMA remain valuable for baseline analysis and interpretability, the empirical evidence from this study suggests that ensemble machine learning techniques provide a more effective and practically relevant solution for transaction-level sales forecasting under dynamic market conditions.

6. Conclusion

This study investigated hourly-level coffee sales forecasting using real transaction data and compared the performance of traditional statistical time series models with machine learning-based approaches. The results show that ensemble machine learning models, particularly XGBoost and Random Forest, substantially outperform the ARIMA model in terms of predictive accuracy, error stability, and explanatory power. Feature importance analysis further indicates that product-specific attributes play a dominant role in determining transaction-level sales, supported by seasonal and intra-day temporal patterns.

The findings confirm the effectiveness of machine learning models in capturing complex, non-linear demand dynamics that traditional time series methods struggle to model. While ARIMA remains useful as a baseline due to its interpretability, its reliance on linear assumptions and aggregated data limits its applicability in highly dynamic retail environments. In contrast, machine learning models demonstrate robustness across varying demand conditions, making them better suited for transaction-level sales forecasting in coffee retail settings.

From a practical perspective, this research provides valuable contributions to coffee retail analytics by enabling more accurate inventory planning, improved staff scheduling, and targeted promotion strategies based on product and temporal insights. Nevertheless, this study is subject to limitations, including reliance on data from a single coffee retailer and the exclusion of external variables such as weather or promotional intensity. Future research should explore multi-store datasets, integrate external and contextual factors, and investigate hybrid or deep learning-based models to further enhance forecasting accuracy and generalizability.

ISSN 2579-7069

7. Declarations

6.1. Author Contributions

Author Contributions: Conceptualization Y.S.W. and A.C.W.; Methodology, Y.S.W. and A.C.W.; Software, Y.S.W.; Validation, Y.S.W.; Formal Analysis, Y.S.W.; Investigation, A.C.W.; Resources, Y.S.W.; Data Curation, A.C.W.; Writing Original Draft Preparation, Y.S.W.; Writing Review and Editing, Y.S.W. and A.C.W.; Visualization, A.C.W. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] P. Ramos and J. M. Oliveira, "Robust Sales Forecasting Using Deep Learning with Static and Dynamic Covariates," *Appl. Syst. Innov*., vol. 6, no. 5, pp. 85, 2023, doi: 10.3390/asi6050085.

[2] E. Obermair, A. Holzapfel, and H. Kühn, "Operational Planning for Public Holidays in Grocery Retailing - managing the Grocery Retail Rush," *Oper. Manag. Res*., vol. 16, no. 2, pp. 931–948, 2023, doi: 10.1007/s12063-022-00342-z.

[3] Y. Zhao, W. Liu, Y. Yan, and F. Li, "Navigating the confluence of econometrics and data science: Implications for economic analysis and policy," *Theor. Nat. Sci*., vol. 38, no. 1, pp. 26–31, 2024, doi: 10.54254/2753-8818/38/20240551.

[4] M. R. Amin, M. Younus, S. Hossen, and A. Rahman, "Enhancing Fashion Forecasting Accuracy Through Consumer Data Analytics: Insights from Current Literature," *Acad. J. Bus. Adm. Innov. Sustain*., vol. 4, no. 2, pp. 54–66, 2024, doi: 10.69593/ajbais.v4i2.69.

[5] A.-F. A. Adetula and T. Akanbi, "Beyond Guesswork: Leveraging AI-driven Predictive Analytics for Enhanced Demand Forecasting and Inventory Optimization in SME Supply Chains," *Int. J. Sci. Res. Arch*., vol. 10, no. 2, pp. 1389–1406, 2023, doi: 10.30574/ijsra.2023.10.2.0988.

[6] N. I. Okeke, O. A. Bakare, and G. O. Achumie, "Artificial Intelligence in SME Financial Decision-Making: Tools for Enhancing Efficiency and Profitability," *Open Access Res. J. Multidiscip. Stud*., vol. 8, no. 1, pp. 150–163, 2024, doi: 10.53022/oarjms.2024.8.1.0056.

[7] G. Theodoridis and A. Tsadiras, "Retail Demand Forecasting: A Comparative Analysis of Deep Neural Networks and the Proposal of LSTMixer, a Linear Model Extension," *Information*, vol. 16, no. 7, pp. 596, 2025, doi: 10.3390/info16070596.

[8] E. Martins and N. V Galegale, "Sales Forecasting Using Machine Learning Algorithms," *Rev. Gestão E Secr*., vol. 14, no. 7, pp. 11294–11308, 2023, doi: 10.7769/gesec.v14i7.1670.

[9] R. Saranya, "Big Mart Sales Predictive Analysis Using Machine Learning," *Int. Sci. J. Eng. Manag*., vol. 02, no. 04, pp. 1–4, 2023, doi: 10.55041/ISJEM00338.

[10] A. Zhu, "VMD-SVM-based Retail Product Demand Forecasting Model," *Sci. J. Econ. Manag. Res*., vol. 7, no. 2, pp. 24–31, 2025, doi: 10.54691/g7pfr887.

ISSN 2579-7069

[11] A. G. Hapsani, "Comparison Various Analytical Approaches to Find the Most Efficient and Effective Method for Peak Hour Identification," *J. Comput. Sci. Inf. Technol*., vol. 17, no. 2, pp. 89–95, 2025, doi: 10.18860/mat.v17i2.29193.

[12] C. Kuo and S. Tsang, "Detection of Price Manipulation Fraud Through Rational Choice Theory: Evidence for the Retail Industry in Taiwan," *Secur. J*., vol. 36, no. 4, pp. 712–731, 2022, doi: 10.1057/s41284-022-00360-3.

[13] T. Berger and J. Koubová, "Forecasting Bitcoin Returns: Econometric Time Series Analysis vs. Machine Learning," *J. Forecast.*, vol. 43, no. 7, pp. 2904–2916, 2024, doi: 10.1002/for.3165.

[14] M. U. Ashraf, "A Predictive Analysis of Retail Sales Forecasting Using Machine Learning Techniques," *Lahore Garrison Univ. Res. J. Comput. Sci. Inf. Technol.*, vol. 6, no. 04, pp. 23–33, 2022, doi: 10.54692/lgurjcsit.2022.0604399.

[15] A. Manoj, A. Baby, and F. S. Devaraj, "Enhancing Enterprise Performance Through Forecasting: A Deep RNN Approach," *Iraqi J. Sci*., vol. 66, no. 10, pp. 4502–4516, 2025, doi: 10.24996/ijs.2025.66.10.36.

[16] A. R. Chowdhury, R. Paul, and F. Z. Rozony, "A Systematic Review Of Demand Forecasting Models for Retail E-Commerce Enhancing Accuracy in Inventory and Delivery Planning," *Int. J. Sci. Interdiscip. Res*., vol. 06, no. 01, pp. 01–27, 2025, doi: 10.63125/mbbfw637.

[17] H.-R. Lou, X. Wang, Y. Gao, and Q. Zeng, "Comparison of ARIMA model, DNN model and LSTM model in predicting disease burden of occupational pneumoconiosis in Tianjin, China," *BMC Public Health*, vol. 22, no. 1, pp. 2167, 2022, doi: 10.1186/s12889-022-14642-3.

[18] Y. Zhao, "E-Commerce Demand Forecasting Using SARIMA Model and K-Means Clustering Analysis," *J. Innov. Dev*., vol. 7, no. 1, pp. 1–6, 2024, doi: 10.54097/ctfb0379.

[19] M. Hadwan, B. M. Al-Maqaleh, F. N. Al-Badani, R. U. Khan, and M. A. Al-Hagery, "A Hybrid Neural Network and Box-Jenkins Models for Time Series Forecasting," *Comput. Mater. Contin*., vol. 70, no. 3, pp. 4829–4845, 2022, doi: 10.32604/cmc.2022.017824.

[20] Y. Chen, X. Zhang, L. Wang, and Z. Li, "Development of a Time Series E-Commerce Sales Prediction Method for Short-Shelf-Life Products Using GRU-LightGBM," *Appl. Sci*., vol. 14, no. 2, pp. 866, 2024, doi: 10.3390/app14020866.

[21] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, "A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks," *Futur. Internet*, vol. 15, no. 8, pp. 255, 2023, doi: 10.3390/fi15080255.

[22] K. T. Kizgin, S. Alp, N. Aydin, and H. Yu, "Machine learning-based sales forecasting during crises: Evidence from a Turkish women's clothing retailer," *Sci. Prog*., vol. 108, no. 1, pp. 1–18, 2025, doi: 10.1177/00368504241307719.

[23] S. Punia, "Medium- to Long-Term Demand Forecasting in Retail and Manufacturing Organizations: Integration of Machine Learning, Human Judgment, and Interval Variable," *J. Forecast*., vol. 45, no. 1, pp. 122–134, 2025, doi: 10.1002/for.70030.

[24] L. Sun, "A Comparative Study of Traditional and Machine Learning Approaches for E-Commerce Sales Forecasting," *Adv. Econ. Manag. Polit. Sci.*, vol. 135, no. 1, pp. 46–51, 2024, doi: 10.54254/2754-1169/2024.18610.

[25] J. Tang, "Unlocking Retail Insights: Predictive Modeling and Customer Segmentation Through Data Analytics," *J. Theor. Appl. Electron. Commer. Res*., vol. 20, no. 2, pp. 59, 2025, doi: 10.3390/jtaer20020059.

[26] T. T. H. Nguyen, A. Bekrar, T. M. Le, M. Abed, and A. Kantasa-Ard, "Toward a Smart Forecasting Model in Supply Chain Management: A Case Study of Coffee in Vietnam," *J. Forecast*., vol. 44, no. 1, pp. 173–199, 2024, doi: 10.1002/for.3189.

[27] T. N. Nguyen, H. P. Tran, M. T. Le, and A. L. Pham, "Product Demand Forecasting with Neural Networks and Macroeconomic Indicators: A Comparative Study Among Product Categories," *J. Bus. Manag. Stud*., vol. 6, no. 2, pp. 170–175, 2024, doi: 10.32996/jbms.2024.6.2.17.