Leveraging Economic and Market Variables for Improved Tile Demand Forecasting using Machine Learning

Sachini Rathnayake[1] and P.P.G. Dinesh Asanka[2]*

[1]Robert Gordon University, Aberdeen, United Kingdom.
[2]Department of Industrial Management, University of Kelaniya, Sri Lanka

*Corresponding author: dasanka@kln.ac.lk

ORCID ID: https://orcid.org/0000-0002-3433-5533

Received: 24.12.2024 Revised: 07.03.2025 Accepted: 17.11.2025 Online: 15.12.2025

Abstract The dynamic nature of the tile market requires precise demand forecasting to ensure inventory optimization, cost control, and enhanced customer satisfaction. This study focuses on developing a robust forecasting model that integrates time series and regression techniques to predict tile sales. Utilizing historical sales data, economic indicators such as GDP growth, inflation, and interest rates, and customer preferences, the research evaluates the performance of various machine learning models, including CatBoost, ARIMA, and Gradient Boosting. The results indicate that models integrating multiple variables significantly outperform those relying on individual factors, with Gradient Boosting and CatBoost demonstrate the highest accuracy for monthly and quarterly forecasts. These findings highlight the value of combining historical sales data, market trends, and economic indicators to achieve reliable sales predictions, thereby enabling more informed decision-making in the tile industry.

Keywords: machine learning, market trends, regression techniques, tile sales forecasting, time series analysis

Introduction

The tile market plays a crucial role in the construction and interior design industries, with its demand intricately linked to economic conditions, customer preferences, and market trends. Accurate demand forecasting in this sector is pivotal for optimizing production, managing inventory, and aligning marketing strategies with customer needs. However, the complexity of predicting sales in a market influenced by fluctuating economic indicators, such as Gross Domestic Production (GDP) growth, inflation, and interest rates, necessitates using advanced analytical techniques (Ahaggach et al., 2024).

This research aims to address these gaps by employing advanced time-series and regression-based machine learning techniques to develop an improved forecasting framework for the tile industry. The key objectives of this study are to analyze historical sales data to identify trends and patterns influencing tile demand across monthly and quarterly time scales, assess the impact of economic indicators such as GDP growth, inflation, and interest rates on sales performance, and incorporate market-specific trends, including customer preferences for tile size, color, and design, into predictive models. By integrating time-series techniques like Seasonal Autoregressive Integrated Moving Average (SARIMA) and Long Short-Term Memory (LSTM) with regression models such as Gradient Boosting and XGBoost (Yan & Liu, 2023), this study evaluates the performance of these methods to identify the most effective approach for forecasting tile sales at different time intervals.

The research problem stems from the inability of traditional forecasting methods to capture the interplay of macroeconomic factors and market-specific variables that influence tile demand. The tile industry, being highly design-sensitive and competitive, requires adaptive forecasting models capable of addressing the challenges posed by its dependence on aesthetics and economic fluctuations. This study addresses this problem by leveraging diverse data sources and advanced analytical techniques, offering actionable insights to optimize inventory levels, improve customer satisfaction, and support data-driven decision-making.

This research makes a significant contribution to the field by providing a comprehensive forecasting framework tailored to the unique challenges of the tile industry. The findings enhance the precision of sales predictions and enable businesses to align their inventory and marketing strategies with anticipated demand, ultimately improving operational efficiency and profitability.

This study investigates the application of time series and regression techniques to enhance the accuracy of tile sales forecasting. By leveraging historical sales data alongside external economic indicators and customer preferences, the research aims to address key challenges in the tile market, including demand volatility and misalignment of supply and demand. The study evaluates multiple machine learning models, including Autoregressive Integrated Moving Average (ARIMA), CatBoost, Gradient Boosting, and LSTM, to identify the most effective monthly and quarterly forecasting approach.

The remainder of this paper discusses the research methodology, analysis of results, and implications for the tile industry. By addressing the limitations of existing forecasting models, this study contributes to the growing body of knowledge on machine learning applications in demand forecasting and offers actionable insights for industry practitioners.

Literature review

This literature review evaluates existing research on sales forecasting in the tile industry, focusing on theoretical frameworks, empirical findings, algorithmic approaches, and research gaps. It explores the application of advanced time-series methods and regression-based machine learning techniques to enhance forecasting accuracy, emphasizing the integration of diverse data sources to address industry-specific challenges.

Sales forecasting in the tile industry relies on foundational theories such as Supply and Demand Theory and Data-Driven Decision-Making (DDDM) Theory. Supply and Demand Theory provides insights into how historical sales data reflects market dynamics, including price elasticity, seasonal trends, and consumer preferences. For instance, Malik et al. (2023) highlighted the relevance of this theory in understanding consumer responses to promotions and seasonal peaks. DDDM Theory emphasizes the importance of leveraging empirical data for strategic decision-making, as demonstrated by Gao and Chen (2022), who showcased the enhanced accuracy of regression-based models like Gradient Boosting Machines (GBM), Cat Boost and XGBoost in capturing demand fluctuations. Economic theories also contribute to understanding how macroeconomic variables such as GDP growth and inflation influence consumer behavior and sector-specific activities. Islam and Amin (2020) demonstrated that inflation dampens consumer spending on non-essential goods like tiles, while Malik et al. (2023) highlighted a strong correlation between GDP growth and increased demand for construction materials.

Historical sales data is a foundational element of demand forecasting, providing insights into recurring patterns, seasonality, and purchasing behaviors. Supply and Demand and Data-Driven Decision-Making theories emphasize how historical data reflects market dynamics and supports strategic decision-making. Empirical studies demonstrate that incorporating machine learning techniques like Random Forest and Neural Networks enhances the predictive power of traditional models. For example, Malik et al. (2023) illustrated how sales data can uncover latent consumer trends, enabling businesses to align supply-side strategies with demand-side patterns.

Economic indicators, including GDP growth, inflation, and interest rates, play a critical role in shaping consumer behavior and market demand. Demand Sensitivity Theory highlights the relationship between macroeconomic variables and discretionary spending, while Macroeconomic Theory provides a broader context by linking aggregate economic trends to sector-specific activities. Research by Malik et al. (2023) and Ranjitha and Spandana (2023) shows how integrating these indicators into regression-based forecasting models significantly improves accuracy. For instance, higher GDP levels are associated with increased construction activity, which drives demand for tiles. Similarly, incorporating inflation metrics helps anticipate demand slowdowns, ensuring businesses remain agile in volatile markets. Market trends, driven by evolving consumer preferences and technological advancements, are another critical dimension. Theories such as Market Trends and Consumer Preferences and Diffusion of Innovation emphasize the dynamic nature of consumer demand. Studies highlight the growing importance of eco-friendly materials, customizable designs, and digital tools like augmented reality in influencing purchasing decisions. For example, Xu and Li (2022) demonstrated how tracking preferences for tile size, color, and design enhances the predictive capability of machine learning models. Integrating these trends into forecasting frameworks allows businesses to respond proactively to shifts in consumer behavior.

Empirical studies reinforce the value of advanced forecasting models in the tile industry. Time-series models such as ARIMA and LSTM have demonstrated significant efficacy in capturing temporal patterns and seasonal trends. Sousa et al. (2023) showed how ARIMA effectively models periodic fluctuations in retail sales, a finding relevant to tile demand influenced by construction cycles. Similarly, Haoudi et al. (2023) highlighted LSTM's ability to capture long-term dependencies, making it ideal for dynamic markets like tiles. Regression-based ensemble models, including GBM,Cat Boost and XGBoost, are particularly effective in handling high-dimensional datasets and non-linear interactions. Falatouri et al. (2023) demonstrated GBM's capability to integrate economic and market variables, achieving superior predictive performance. Kheawpeam and Sinthupinyo (2023) also validated XGBoost's ability to account for complex interactions between macroeconomic factors and customer preferences, making it highly suitable for tile sales forecasting. The integration of diverse data sources, such as historical sales records, macroeconomic indicators, and market trends, enhances predictive accuracy. Van Nguyen et al. (2020) demonstrated the benefits of integrating distributor data with customer demographics and pricing trends in retail sales forecasting, an approach that is directly applicable to the tile industry.

Several algorithms have been critically reviewed for their applicability in predicting tile industry demand. Random Forest, an ensemble learning method, has demonstrated robustness in managing non-linear relationships and high-dimensional datasets, as shown by Gao and Chen (2022). However, its computational intensity and lack of interpretability pose challenges for large-scale applications. Decision Trees offer simplicity and interpretability, making them effective for identifying relationships between promotional activities and sales, as Pereira and Cerqueira (2022) highlighted. However, they are prone to overfitting and instability. Gradient Boosting Machines (GBM) and XGBoost, widely used regression models, excel in integrating diverse variables and achieving high predictive accuracy. Falatouri et al. (2023) emphasized GBM's superior performance in reducing error rates, while Kheawpeam (2023) demonstrated XGBoost's scalability and efficiency in handling high-dimensional data. SARIMAX, a time-series model, effectively incorporates exogenous variables such as construction activity and economic trends. Sousa et al. (2023) demonstrated its suitability for industries with seasonal demand patterns, like tiles. LSTM networks, designed for sequential data, excel in capturing long-term dependencies and patterns. Haoudi et al. (2023) highlighted LSTM's ability to adapt to evolving market trends, although its computational intensity and lack of interpretability remain limitations.

Despite these advancements, several gaps remain in the application of forecasting models to the tile industry. The limited use of time-series and regression models in this sector restricts the ability to capture temporal dependencies and seasonal trends unique to tile sales. Macroeconomic indicators, such as GDP growth, inflation, and interest rates, are often excluded from tile sales forecasting models, leading to incomplete predictions. Xu and Li (2022) emphasized the importance of integrating dynamic market trends, such as consumer preferences for tile size, color, and design, into predictive models. However, existing models frequently focus solely on historical sales data. The lack of hybrid modeling approaches, which combine time-series and regression techniques, limits the ability to simultaneously capture non-linear interactions and temporal dependencies. Studies by Zhang et al. (2023) and Kheawpeam and Sinthupinyo (2023) demonstrated the success of hybrid models in other sectors, but their adoption in the tile industry remains minimal. Additionally, the failure to integrate diverse data sources, such as macroeconomic indicators, market trends, and historical sales records, reduces the robustness and effectiveness of existing models.

This study addresses these gaps by developing a hybrid forecasting framework that integrates advanced time-series and regression-based machine learning techniques. This research aims to enhance predictive accuracy and provide actionable insights for optimizing inventory management, marketing strategies, and operational efficiency in the tile industry by leveraging historical sales data, macroeconomic indicators, and market trends.

Methodology

This research employs a comprehensive quantitative approach to develop and evaluate machine learning models for accurately forecasting sales in the tile industry. The methodology is structured to integrate diverse data sources—historical sales, macroeconomic indicators, and market trends—and optimize predictive accuracy through advanced analytical techniques. The key stages of the methodology include research design, data collection, preprocessing, feature engineering, model training, and evaluation (Figure 1).

The research adopts a quantitative research design focusing on the systematic analysis of numerical data to uncover patterns and relationships within the tile market. Data is collected from multiple sources, including ERP systems, government publications, and market trend reports. The dataset spans from April 2022 to March 2024, incorporating ERP system records, government and industry reports, and market trend data on consumer preferences. Historical sales data provides granular insights into past transaction volumes, revenue, and customer segmentation. Macroeconomic indicators such as GDP growth, inflation rates, and interest rates capture external influences on consumer behavior. Market trends, including tile size, color, and design preferences, are included to reflect dynamic consumer preferences.

Data preprocessing ensures the quality and compatibility of the dataset. Advanced techniques impute missing values, and data normalization aligns scales across variables. Feature engineering transforms raw data into actionable inputs for modeling by encoding categorical variables, creating lagged variables for time-series analysis, and decomposing seasonal trends. These steps enhance the dataset's ability to capture temporal and contextual patterns in sales data.

Initially, missing values were handled by implementing appropriate imputation techniques. For numerical variables such as Tiles SQM, Net Sales, and Gross Sales, missing values were replaced using the mean of the respective columns. In contrast, categorical variables including Item Size, Collection, District, Province, Company, and Sub Category 1 were filled using the mode, ensuring that missing categorical entries were replaced with the most frequently occurring value.

To maintain data integrity, an outlier detection and removal process was applied using the Z-score method. Extreme values in numerical variables such as Net Sales, Gross Sales, and Tiles SQM were removed based on a Z-score threshold of 3, leading to the elimination of 4,337 outliers. This process ensured that anomalies did not adversely impact the performance of the forecasting models.

Categorical variables were encoded to facilitate machine learning model training. Label encoding was used for ordinal categorical variables where order was meaningful, while one-hot encoding was applied to nominal categorical variables, converting them into binary features to maintain model interpretability. Furthermore, time-series structuring was incorporated into the dataset, ensuring chronological consistency for forecasting. Lag variables were created to capture historical patterns, and seasonal decomposition was performed to isolate trend, seasonal, and residual components.

Since time-series forecasting requires stationary data, the Augmented Dickey-Fuller (ADF) test was conducted on the Gross Sales variable to check for stationarity. The null hypothesis (H_0) assumes the presence of a unit root, indicating a non-stationary series, while the alternative hypothesis (H_1) suggests a stationary series. The test results yielded a test statistic of -37.1682 with a p-value of 0.0000, alongside critical values at different confidence levels: -3.4303 (1%), -2.8615 (5%), and -2.5667 (10%). Since the test statistic is significantly lower than all critical values and the p-value is less than 0.05, the null hypothesis is rejected, confirming that the GROSS SALES time-series is stationary. This means the data does not require differencing before applying forecasting models, ensuring more accurate predictions.

The method of sampling involves purposive sampling of secondary data from the company's ERP system and external publications. This non-probability sampling technique ensures the inclusion of datasets that are directly relevant to the research objectives, such as historical sales, economic indicators, and market trends. The two-year data period provides a balance between recency and sufficient historical context, aligning with best practices in demand forecasting literature. This study aims to develop a comprehensive and reliable forecasting framework for the tile industry by combining targeted data collection with advanced analytical techniques.

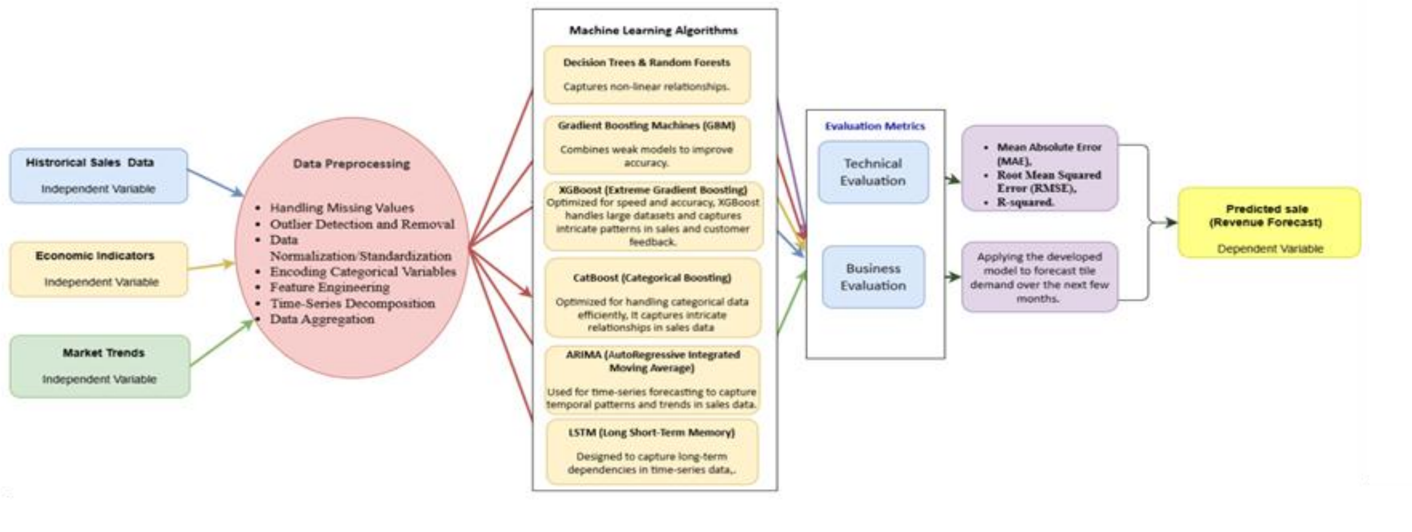

Figure 1: Methodology

As outlined in Figure 1. Machine learning models are developed to capture intricate relationships within the dataset. Decision Trees are used to segment sales data based on factors such as product size, seasonality, and consumer preferences. Random Forest and Gradient Boosting Machines (GBM) are implemented to manage nonlinear relationships and identify key variables influencing sales. XGBoost, an advanced implementation of GBM, ensures scalability and reduces overfitting through regularization. CatBoost is incorporated for handling categorical variables efficiently while maintaining high accuracy in scenarios with high-cardinality features. ARIMA is used to model temporal patterns, incorporating exogenous macroeconomic variables to capture trends and seasonality, while Long Short-Term Memory (LSTM) networks analyze sequential dependencies and long-term temporal patterns in the data.

The dataset is split into training and testing subsets to ensure model reliability using the 80/20 rule. The training dataset consists of 80% of the total data and is used to train the machine learning models, while the remaining 20% is reserved for testing and evaluating the models' predictive performance. The split is implemented chronologically to prevent data leakage and maintain the integrity of the time-series patterns, allowing the models to generalize effectively across unseen data.

The evaluation of these models incorporates both technical and business metrics. Technical metrics include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared to assess predictive accuracy. Business metrics focus on operational impacts, such as optimizing inventory management, aligning marketing strategies, and improving supply chain efficiency.

As shown in Figure 2. the conceptual framework integrates three key independent variables—historical sales data, economic indicators, and market trends—to predict sales revenue (dependent variable) in the tile market using time-series and regression techniques. This holistic approach addresses the complexities of forecasting by incorporating both internal and external factors. Historical sales data provides a foundation for understanding past performance, capturing seasonality, and identifying recurring patterns. Research confirms its importance in predicting future trends by modeling temporal dependencies (Sousa et al., 2023). Economic indicators, such as GDP growth, inflation, and interest rates, reflect macroeconomic conditions that influence consumer spending and construction activity. These variables are critical for explaining variations in tile demand driven by broader economic shifts (Wang & James, 2024). Market trends, including consumer preferences for tile size, colour, and design, capture the dynamic nature of customer behavior and aesthetic shifts in the industry, ensuring the forecasting model remains adaptive to market changes (Chae et al., 2024).

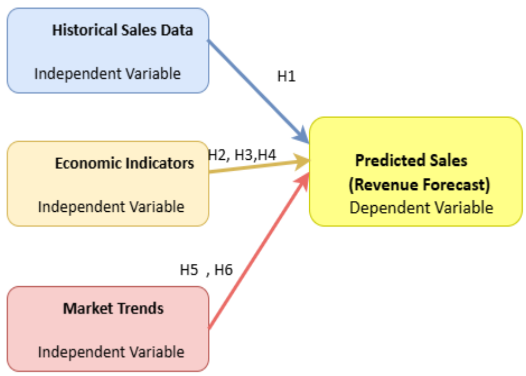

Figure 2: Conceptual Framework

This research proposes several hypotheses to explore the relationship between the independent variables and the dependent variable (predicted sales revenue). Hypothesis H1 posits that historical sales data positively impacts revenue predictions. Hypotheses H2, H3, and H4 evaluate the effects of GDP growth, inflation, and interest rates on tile sales. Finally, H5 and H6 assess how market trends influence sales performance, such as preferences for larger tile sizes and light-colored designs. These hypotheses aim to validate the model's ability to incorporate diverse factors for enhanced accuracy.

As shown in Figure 3. the operationalization of variables ensures clarity and consistency in measurement. Historical sales data includes metrics like transaction dates, sales volume, and product categories, measured on continuous and nominal scales. Economic indicators, such as GDP growth, inflation, and interest rates, are quantified using continuous scales. Market trends are represented through dimensions like tile color, size, and design preferences, operationalized via ordinal and nominal scales. The dependent variable, revenue forecast, is evaluated using predictive metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared, ensuring the reliability and validity of the forecasting model.

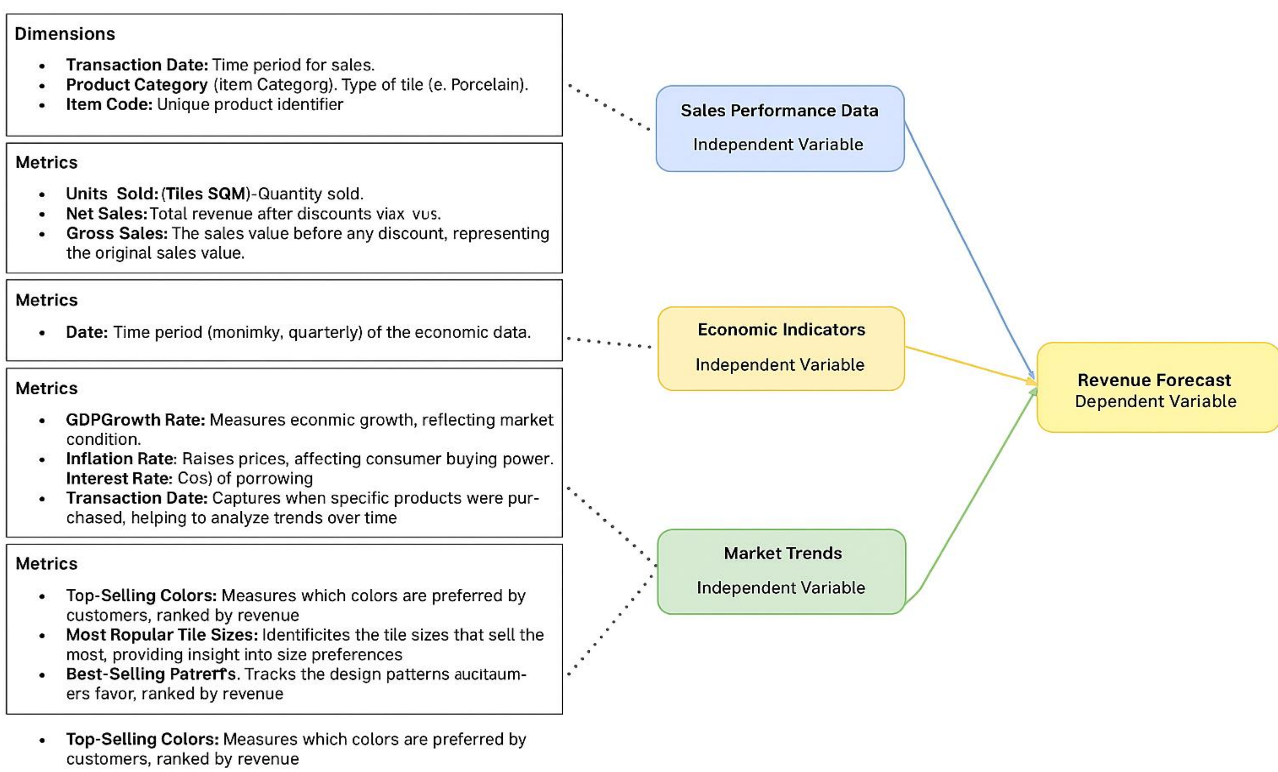

Figure 3: Operationalization of variables for forecasting of Tile Sales

Developing tile sales forecasting model

Defining Hypothesis on Tile Sales

This research proposes several hypotheses to explore the relationship between the independent variables and the dependent variable (predicted sales revenue). Hypothesis H1 posits that historical sales data positively impacts revenue predictions. Hypotheses H2, H3, and H4 evaluate the effects of GDP growth, inflation, and interest rates on tile sales. Finally, H5 and H6 assess how market trends influence sales performance, such as preferences for larger tile sizes and light-colored designs. These hypotheses aim to validate the model's ability to incorporate diverse factors for enhanced accuracy.

These hypotheses, aligned with H1 through H6, were rigorously tested on both monthly and quarterly data to understand short-term and long-term trends. Each hypothesis was paired with advanced machine learning algorithms, whose performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared metrics. The hypotheses are as follows.

- H1: Historical sales data has a significant positive relationship with predicted sales revenue.
- H_o1: Historical sales data does not have a significant positive relationship with predicted sales revenue.
- H2: GDP growth has a significant positive impact on predicted sales revenue in the tile market.
- H_o2: GDP growth does not significantly impact predicted sales revenue in the tile market.
- H3: Inflation negatively impacts predicted sales revenue in the tile market.
- H_o3: Inflation does not negatively impact predicted sales revenue in the tile market.
- H4: Interest rates negatively impact predicted sales revenue in the tile market.
- H_o4: Interest rates do not negatively impact predicted sales revenue in the tile market.
- H5: Customer preference for larger tile sizes is not positively correlated with increased tile sales revenue..
- H_o5: Customer preference for larger tile sizes is positively correlated with increased tile sales revenue..
- H6: Increasing demand for light-colored tiles (e.g., white, beige) is positively associated with overall tile sales revenue.

Performing Hypothesis Testing for Monthly Sales

Monthly analysis provides insights into short-term trends and relationships. For H1, historical sales data shows consistently strong positive correlations with gross sales, with net sales and tile quantities sold demonstrating correlations above 0.92 (p-value $<$ 0.05) across multiple months. This confirms the foundational role of historical sales data in monthly forecasting.

For H2:((GDP Growth),The correlation between GDP growth and gross sales was inconsistent on a monthly basis, ranging from -0.10 in February ($p < 0.001$) to 0.03 in May ($p < 0.05$). While GDP growth showed weak to negligible relationships with sales revenue monthly, certain months, such as April, revealed slightly positive correlations (0.02, $p < 0.01$). Similarly, H3 (inflation) and H4 (interest rates) exhibit mixed and weak correlations every month, ranging from -0.08 to 0.11, indicating limited predictive utility at this temporal resolution. H3 (inflation) exhibited both positive and negative correlations with gross sales across months. For example, positive correlations such as 0.10 in February ($p < 0.001$) and 0.07 in January ($p < 0.001$) were observed, but negative correlations appeared in months like December (-0.03, $p < 0.001$). The findings suggest that the impact of inflation on sales revenue varies by month, influenced by external market conditions. For H4(Interest Rates), Monthly correlations between interest rates and gross sales showed minimal or negative relationships, with significant values like -0.12 in July ($p < 0.001$) and -0.08 in June ($p < 0.001$). These results confirm that higher interest rates can deter customer spending on tiles.

H5 (tile size preferences) and H6 (light-colored tile preferences) exhibit consistently strong positive correlations for market trends. Larger tile sizes demonstrated strong positive correlations with sales revenue across months. For instance, 600 x 600 mm tiles consistently showed correlations exceeding 0.92 ($p < 0.001$) throughout various months, confirming their popularity and significant contribution to revenue and Light-colored tiles, such as white and beige, exhibited high positive correlations with sales revenue, ranging from 0.91 to 0.94 ($p < 0.001$) across months like January and August. This emphasizes the increasing demand for aesthetic and modern tile preferences.

Performing Hypothesis Testing for Quarterly Sales

Quarterly testing reveals stronger and more consistent relationships for economic indicators. For H2 (GDP growth), correlations with gross sales are significantly positive across all quarters, ranging from 0.94 to 0.96 (p-value < 0.05). This highlights the importance of aggregating data quarterly to capture macroeconomic effects. Conversely, H3 (inflation) and H4 (interest rates) demonstrate consistently negative correlations with gross sales (inflation: -0.63 to -0.67, interest rates: -0.50 to -0.66, p-value < 0.05), suggesting that these factors exert a stronger influence over longer temporal intervals.

For H5 and H6, preferences for larger tile sizes (e.g., 600 x 600 mm) and light-colored tiles (e.g., white, beige) remain robustly positive on a quarterly basis, with correlations exceeding 0.9 (p-value < 0.05). These findings indicate that market trends are critical for sales forecasting regardless of the temporal aggregation.

The hypothesis testing results suggest that historical sales data and market trends are effective predictors on both monthly and quarterly scales. However, economic indicators such as GDP growth, inflation, and interest rates exhibit stronger and more reliable correlations at a quarterly level, aligning with the broader economic cycles that influence consumer behavior. These insights underscore the importance of incorporating economic factors on a quarterly basis to enhance the robustness and accuracy of predictive models for tile sales forecasting. By integrating these diverse data sources, the study confirms the overall hypothesis that a combined approach improves forecasting outcomes significantly. Comparison of Hypotheses (H1 – H6) for Monthly vs Quarterly shown in the Table 1.

Table 1. Monthly vs. Quarterly Correlation Summary for Tile Sales Forecasting Hypotheses (H1–H6)

Hypothesis	Monthly Sales Correlation Summary	Quarterly Sales Correlation Summary
H1 – Historical Sales Data	Strong positive correlations between historical sales, net sales, and quantities sold (> 0.92, $p < 0.05$) across all months. Confirms the foundational role of historical data in short-term forecasting.	Maintains strong positive relationships across quarters (> 0.94, $p < 0.05$), reaffirming the dominant predictive power of historical sales data.
H2 – GDP Growth	Inconsistent correlations, ranging from −0.10 (Feb, $p < 0.001$) to 0.03 (May, $p < 0.05$). Weak or negligible impact monthly.	Strong positive correlations with gross sales (0.94 – 0.96, $p < 0.05$), showing significant influence when aggregated quarterly.
H3 – Inflation	Mixed correlations (−0.03 to 0.10). Positive in some months (e.g., Jan = 0.07, Feb = 0.10; $p < 0.001$), negative in others. Indicates variable short-term impact.	Consistently negative correlations with gross sales (−0.63 to −0.67, $p < 0.05$), suggesting inflation exerts stronger influence over longer time frames.
H4 – Interest Rates	Mostly negative or weak correlations (e.g., Jun = −0.08, Jul = −0.12; $p < 0.001$). Higher rates deter tile purchases.	Stronger negative correlations (−0.50 to −0.66, $p < 0.05$), confirming that interest rates significantly reduce sales on a quarterly basis.
H5 – Tile Size Preferences	Strong positive correlations between large tile sizes (e.g., 600 × 600 mm) and sales (> 0.92, $p < 0.001$). Demonstrates consistent monthly preference for larger tiles.	Robust positive correlations (> 0.90, $p < 0.05$). Market trend remains stable and predictive across quarters.
H6 – Light-Colored Tile Preferences	Strong positive correlations (0.91 – 0.94, $p < 0.001$) for light-colored tiles (white, beige) across several months, reflecting aesthetic consumer preferences.	Consistently strong positive correlations (> 0.90, $p < 0.05$), confirming enduring demand for light-colored, modern designs.

Tile Sales Forecasting Model Performance

The performance of monthly sales forecasting model is summarized shown in Table 2. The evaluation of models for monthly sales forecasting reaffirms Gradient Boosting as the most reliable model, achieving the highest R-squared value of 0.97566, a low Mean Absolute Error (MAE) of 17,230.0, and a Root Mean Squared Error (RMSE) of 22,803.1. However, it is noticeable that most of the model have resulted similar values among the evaluation parameters. These metrics demonstrate Gradient Boosting's superior ability to model non-linear relationships and capture short-term fluctuations in sales data. This capability is particularly crucial in the dynamic tile industry, where seasonality, promotions, and customer preferences influence monthly sales. Random Forest and CatBoost also demonstrated strong performance, achieving R-squared values of 0.97285 and 0.97307, respectively, with comparable error metrics. However, both models slightly lagged behind Gradient Boosting in precision and overall accuracy, solidifying Gradient Boosting's position as the best model for monthly forecasts. Conversely, ARIMA exhibited significant instability in monthly forecasting, with the highest MAE of 37,587.9, the highest RMSE of 48,392.3, and the lowest R-squared value of 0.90080. These results underscore ARIMA's inability to capture the complex, non-linear relationships inherent in monthly sales data, likely due to its linear time-series framework. This underperformance highlights ARIMA's limitations in capturing non-linear relationships inherent in monthly sales data. Additionally, Extreme Gradient Boosting (XGBoost) recorded higher error metrics (MAE: 17,473.5, RMSE: 24,674.8) than Gradient Boosting and CatBoost, suggesting it is less favorable for monthly forecasting. Overall, Gradient Boosting is identified as the optimal model for monthly predictions due to its balance of high accuracy, robustness, and effective error minimization.

Table 2. Tiles Sales Monthly Forecasting Model Performance Comparison

Model	Performance Metrics		
	Mean Absolute Error (MAE)	***Root Mean Squared Error (RMSE)***	***R Squared***
Decision Trees	17,110.50	27,869.40	0.963
Random Forest	14,652.60	24,544.70	0.971
Gradient Boosting	14,241.90	23,224.20	0.975
Extreme Gradient Boosting	14,349.90	23,368.60	0.974
LSTM	14,960.70	23,918.00	0.973
ARIMA	18,717.20	30,191.50	0.961
Cat Boosting	**13,567.90**	22,926.00	0.975

Based on our research findings, the monthly tile sales can be predicted as follows:

$$\hat{Y}_{monthly} = \sum_{m=1}^{M} \eta . h_m(X) + \varepsilon \; EQ(1)$$

where $\hat{Y}_{monthly}$ is the predicted sales revenue, M is the number of trees, η is the learning rate, $h_m(X)$ is the output of the m^{th} tree, and ϵ is the residual error. The performance of quarterly sales predictive model is summarized shown in Table 1. CatBoost emerges as the top-performing model for quarterly sales forecasting, achieving the lowest MAE of 13,567.9, the lowest RMSE of 22,926.0, and a high R-squared value of 0.97540. CatBoost's specialized handling of categorical variables and its efficiency in modeling datasets with high dimensionality make it particularly adept at capturing macroeconomic trends and long-term market dynamics that drive quarterly sales. Gradient Boosting also performed well, achieving an R-squared value of 0.97475; however, it was slightly less effective than CatBoost in minimizing error metrics, with a higher MAE of 14,241.9 and RMSE of 23,224.2. Random Forest also exhibited competitive performance, achieving an R-squared of 0.97180, MAE of 14,652.6, and RMSE of 24,544.7, positioning it as a viable alternative.

In contrast, ARIMA demonstrated instability even in quarterly forecasting. Despite an improvement over its monthly results, ARIMA's R-squared value of 0.96139, MAE of 18,717.2, and RMSE of 30,191.5 reflect its limitations in capturing aggregated trends and the macroeconomic complexities necessary for accurate quarterly predictions. This underperformance likely stems from ARIMA's reliance on linearity and its sensitivity to external disruptions. Based on these results, CatBoost is recommended as the best model for quarterly forecasting, delivering superior accuracy and effectively capturing the nuanced, long-term dynamics of the tile industry.

Table 3. Tiles Sales Quarterly Forecasting Model Performance Comparison

Model	Performance Metrics		
	Mean Absolute Error (MAE)	***Root Mean Squared Error (RMSE)***	***R Squared***
Decision Trees	20,435.50	30,954.90	0.955
Random Forest	16,359.90	24,084.70	0.972
Gradient Boosting	17,230.00	22,803.10	0.976
Gradient Boosting	14,241.90	23,224.20	0.975
Extreme Gradient Boosting	17,473.50	24,674.80	0.972
LSTM	16,929.30	23,013.20	0.975
ARIMA	37,587.90	48,392.30	0.901
Cat Boosting	16,067.10	23,986.70	0.973

Based on our research findings, the quarterly tile sales can be predicted as follows:,

$$\hat{Y}_{quarterly} = \sum_{i=1}^{N} f_i(X_{cat}, X_{num}) + \varepsilon \quad EQ(2)$$

Where $\hat{Y}_{quarterly}$ is the predicted sales revenue, N is the number of features, $f_i(X_{cat}, X_{num})$ represents the contribution of categorical (X_{cat}) and numerical (X_{num}) variables, and ϵ is the residual error.

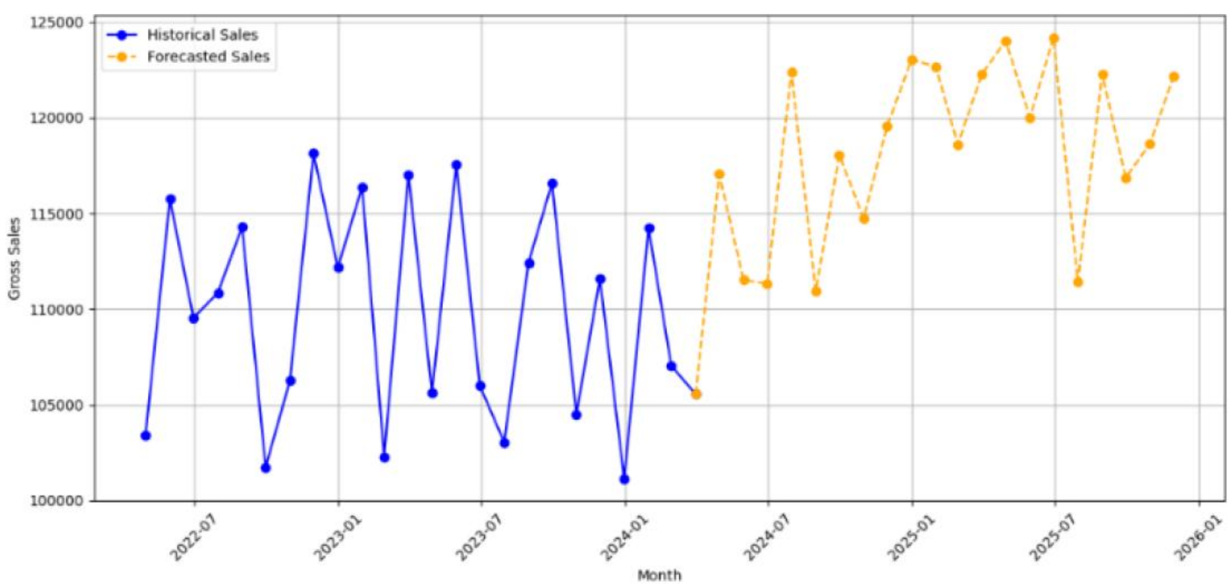

Figure 4: Predicted monthly tile sales using the proposed Gradient Boosting model

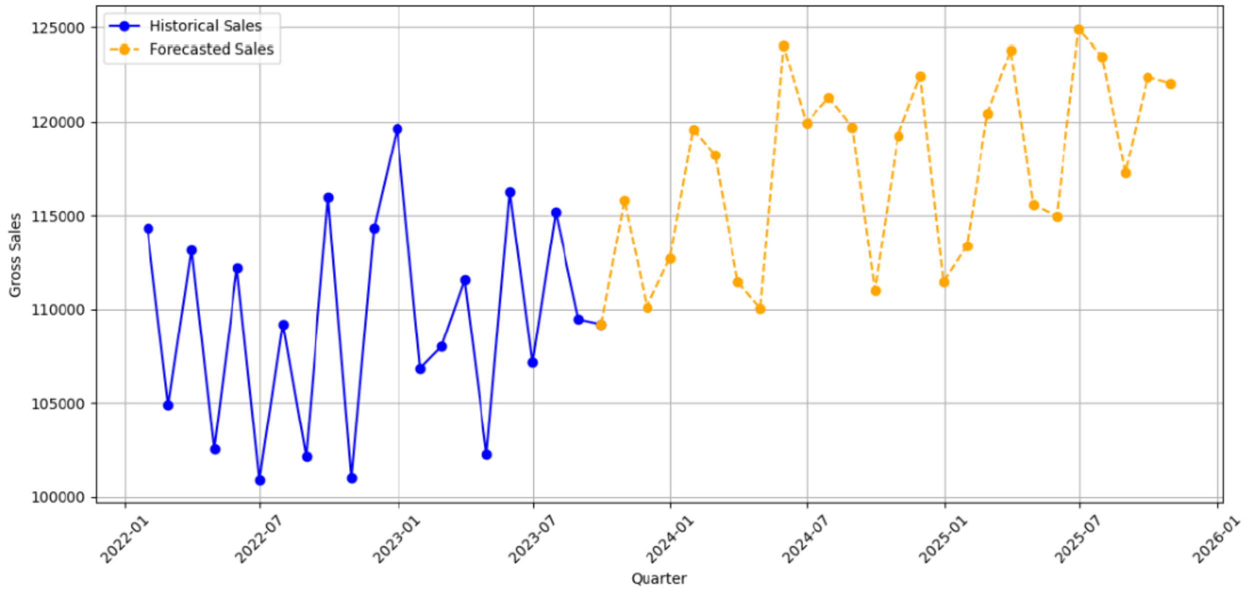

Figure 5: Predicted quarterly tile sales using the proposed Cat Boost model

Recommendations to Enhance Business Strategies

As shown in Figure 4 and Figure 5, the monthly and quarterly sales forecasts generated using Gradient Boosting and CatBoost models provide valuable insights into historical trends and future demand patterns in the tile market. These visual representations demonstrate how predictive modeling enhances business decision-making across multiple operational areas, including inventory management, financial planning, marketing strategy, and risk mitigation. By analyzing both short-term (monthly) and long-term (quarterly) forecasts, businesses can develop data-driven strategies that optimize performance and maintain competitive advantages.

Conversely, the quarterly sales forecast (Figure 4) provides a broader, long-term perspective on demand patterns, highlighting stable trends across multiple quarters. The historical sales trend (blue) shows volatility across previous quarters, while the forecasted trend (orange) projects a more structured pattern of growth. This longer forecasting window allows supply chain managers to plan bulk procurement, optimize logistics, and manage supplier contracts more efficiently. The visibility into quarterly trends also enables better warehouse capacity planning, ensuring businesses can meet projected demand without overcommitting resources.

Inventory Management and Supply Chain Optimization

The monthly sales forecast (Figure 3) highlights fluctuations in demand at a granular level, allowing businesses to identify seasonal trends and short-term market shifts. The historical sales trends, represented by the blue line, indicate significant monthly variations in sales volumes, reflecting the impact of factors such as promotional campaigns, seasonal demand, and consumer purchasing behavior. The forecasted sales trend (orange line) continues this pattern, accurately estimating expected future sales. This insight enables inventory managers to align procurement cycles and stock levels accordingly, preventing excess inventory (which leads to storage costs) and stockouts (resulting in lost sales).

Financial Planning and Revenue Management

Both monthly and quarterly forecasts play a crucial role in financial stability and revenue forecasting. The monthly forecast (Figure 3) provides detailed, short-term projections that assist businesses in managing cash flow, budgeting, and planning operational expenditures. The ability to anticipate fluctuations in sales allows financial planners to allocate resources efficiently for ongoing operational costs, including marketing, staffing, and production.

The quarterly forecast (Figure 4) offers macro-level financial insights, helping businesses predict revenue trends for a more extended period. The upward trajectory in the forecasted sales trend suggests continued demand growth, reinforcing the importance of long-term investment decisions. Businesses can use these insights to expand production capacity, explore new market opportunities, or allocate funds for strategic business initiatives such as technological upgrades or product diversification.

Marketing Strategy Development and Sales Optimization

The sales trends depicted in Figure 3 and Figure 4 are essential for refining marketing strategies and optimizing sales performance. The monthly trend analysis reveals sharp peaks and dips in historical sales, indicating the influence of marketing campaigns, promotions, and seasonal shopping patterns. By analyzing these fluctuations, businesses can plan timely advertising efforts, launch discount offers during low-demand months, and capitalize on high-demand periods with targeted promotions.

The quarterly forecast (Figure 4), on the other hand, provides a strategic view of market expansion opportunities. The stable, forecasted sales trend suggests consistent consumer demand, making it an ideal reference point for long-term marketing initiatives. Businesses can align promotional activities with forecasted demand surges, ensuring that campaigns are rolled out when consumer interest is at its peak. Furthermore, quarterly analysis allows for regional demand segmentation, enabling businesses to prioritize specific markets based on projected growth rates.

Risk Mitigation and Business Continuity Planning

Figure 3. and Figure 4. demonstrate that demand forecasting plays a crucial role in identifying potential risks and implementing proactive business continuity strategies. The monthly trend (Figure 3) highlights short-term demand volatility, indicating potential risks associated with unexpected market fluctuations, economic shifts, or competitive disruptions. Businesses can use these insights to adjust pricing strategies, diversify product offerings, and implement dynamic sales models that ensure financial resilience.

The quarterly forecast (Figure 4) reflects broader macroeconomic influences on the tile market. The inclusion of economic indicators such as GDP growth, inflation, and interest rates within the forecasting models ensures that businesses are prepared for long-term financial uncertainties. For example, a decline in forecasted sales over consecutive quarters may indicate economic downturns, shifts in consumer spending habits, or changes in interest rates, allowing companies to adjust their capital investments, renegotiate supplier contracts, and modify expansion plans accordingly.

Strategic Decision-Making for Business Growth

The forecasting models also support long-term strategic planning by offering predictive insights into market demand. The gradual increase in forecasted sales in Figure 4 suggests a growing market trend, which businesses can leverage to expand product lines, enter new markets, and explore investment opportunities. Additionally, the monthly forecast (Figure 3) provides critical insights into consumer preferences for specific tile colors, sizes, and designs, allowing businesses to tailor their product offerings to match evolving trends.

By integrating Gradient Boosting and CatBoost models into their forecasting frameworks, businesses in the tile industry can move beyond traditional, reactive decision-making and adopt a data-driven, predictive approach. This transition enables firms to stay ahead of market trends, optimize operational efficiency, and maintain a competitive advantage in an increasingly dynamic industry.

Conclusion

This research aimed to enhance tile sales forecasting by integrating historical sales data, economic indicators, and market trends using advanced machine learning algorithms. The study rigorously tested hypotheses to evaluate the significance of these variables on sales revenue, examining both short-term (monthly) and long-term (quarterly) trends.

The findings confirmed that historical sales data and market trends, such as customer preferences for larger tile sizes and light-colored tiles, are strong predictors of sales performance across both temporal resolutions. These variables consistently demonstrated high correlations with sales revenue, reinforcing their foundational role in demand forecasting. On the other hand, economic indicators, including GDP growth, inflation, and interest rates, exhibited stronger and more consistent relationships with sales revenue on a quarterly basis, aligning with broader economic cycles and their influence on consumer behavior.

From a modeling perspective, Gradient Boosting emerged as the most reliable approach for monthly forecasting, effectively capturing non-linear relationships and short-term fluctuations. Its high R-squared value and minimal error metrics made it the optimal choice for handling the dynamic nature of monthly sales. In contrast, CatBoost excelled in quarterly forecasting, demonstrating superior accuracy and robustness in modeling aggregated trends and macroeconomic dynamics. The study also highlighted the limitations of traditional methods like ARIMA, which struggled to address the non-linear and complex relationships inherent in sales data, rendering it less suitable for both monthly and quarterly forecasting.

The research validates the overarching hypothesis that combining historical sales data, economic indicators, and market trends significantly enhances forecasting accuracy compared to relying on individual variables. Integrating these diverse data sources allows businesses in the tile industry to predict demand with greater precision, supporting strategic decision-making in inventory management, marketing, and production planning.

In conclusion, this study provides actionable insights such as inventory optimization, dynamic pricing strategies, targeted marketing and production scheduling etc. for practitioners and researchers by identifying the most effective forecasting models for different temporal needs. Gradient Boosting and CatBoost stand out as the best monthly and quarterly predictions models, respectively, ensuring reliable and accurate sales forecasts. Future research could further refine these models by incorporating real-time data and exploring advanced deep learning techniques, such as hybrid models, to address evolving market dynamics. This would further strengthen the industry's ability to adapt to rapidly changing consumer preferences and macroeconomic conditions.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Author Contributions

Sachini Rathnayake contributed to the development ofthe review concept, conducted the literature search, data collection and data analysis, and prepared the original manuscript draft. PPG Dinesh Asanka provided expert guidance on the scientific aspects related to Time Series forecasting, refined the scope of the review, contributed ideas for content development, and reviewed and edited

References

Ahaggach, H., Abrouk, L., & Lebon, É. (2024). Systematic Mapping Study of Sales Forecasting: Methods, Trends, and Future Directions. *Forecasting,* 6(3), 502–532. https://doi.org/10.3390/forecast6030028

Alapati, N., Prasad, B. V. V. S., & Sharma, A. (2022). Prediction of flight fare using machine learning. *Proceedings of the International Conference on Future Innovations in Research, Technology, and Practice* (pp. 100–108). https://doi.org/10.1109/ICFIRTP56122.2022.10059429

Aichner, T., & Santa, V. (2023). Demand forecasting methods and the potential of machine learning in the FMCG retail industry. In Serving the Customer: The Role of Selling and Sales. *Research Papers in Economics* (pp. 215–252). https://doi.org/10.1007/978-3-658-39072-3

Antunes, A., Andrade-Campos, A., Sardinha-Lourenço, A., & Oliveira, M. S. (2018). Short-term water demand forecasting using machine learning techniques. *Journal of Hydroinformatics*. https://doi.org/10.2166/hydro.2018.163

Chae, B. K., Sheu, C., & Park, E. O. (2024). The value of data, machine learning, and deep learning in restaurant demand forecasting: Insights and lessons learned from a large restaurant chain. *Decision Support Systems*, 184, Article 114291. https://doi.org/10.1016/j.dss.2024.114291

Do, L. P. C., Hagfors, L. I., Lin, K.-H., & Molnár, P. (2016). Demand and residual demand modeling using quantile regression. *Renewable Energy and Environmental Sustainability*, 1, Article 41. https://doi.org/10.1051/rees/2016045

Falatouri, T., Brandtner, P., & Darbanian, F. (2023). Applying machine learning in retail demand prediction—A comparison of tree-based ensembles and long short-term memory-based deep learning. *Applied Sciences*, 13(19), Article 11112. https://doi.org/10.3390/app13 1911112

Eiglsperger, J., Haselbeck, F., Stiele, V., Guadarrama Serrano, C., Lim-Trinh, K., Menrad, K., Hannus, T., & Grimm, D. G. (2024). Forecasting seasonally fluctuating sales of perishable products in the horticultural industry. *Expert Systems With Applications*, 249, Article 123438. https://doi.org/10.1016/j.eswa.2024.123438

Gao, C., & Chen, Y. (2022). Using machine learning methods to predict demand for bike sharing. *Proceedings of the ENTER 2022 eTourism Conference*. https://doi.org/10.1007/978-3-030-94751-4

Geetha Mani, G., Viswanadhapalli, J. K., & Stonier, A. A. (2021). Prediction and forecasting of air quality index in Chennai using regression and ARIMA time series models. Journal of Engineering Research, 10(2A), 179–194. https://doi.org/10.36909/jer.10253

Golderzahi, V., & Pao, H.-K. (2024). Revenue forecasting in smart retail based on customer clustering analysis. *Internet of Things*, 27, Article 101286. https://doi.org/10.1016/j.iot.202 4.101286

Gustriansyah, R., Ermatita, E., & Rini, D. P. (2022). An approach for sales forecasting. *Expert Systems With Applications*, 207, Article 118043. https://doi.org/10.1016/j.eswa.2022.118043

Haoudi, Y., Yazdani, M. A., Roy, D., & Hennequin, S. (2023). Demand prediction based on machine learning algorithms for optimal distribution of insulin. *IFAC PapersOnLine*, 56(2), 10174–10179. https://doi.org/10.1016/j.ifacol.2023.10.893

Huber, J., & Stuckenschmidt, H. (2020). Daily retail demand forecasting using machine learning with emphasis on calendric special days. *International Journal of Forecasting*, 36, 1420–1438. https://doi.org/10.1016/j.ijforecast.2020.02.005

Islam, S., & Amin, S. H. (2020). Prediction of probable backorder scenarios in the supply chain using distributed random forest and gradient boosting machine learning techniques. *Journal of Big Data*, 7(65), 1–22. https://doi.org/10.1186/s40537-020-00345-2

Jayapal, K. (2022). Food demand prediction using statistical and machine learning models (Master's thesis). National College of Ireland.

Karunanithi, M., Chatasawapreeda, P., & Khan, T. A. (2024). Predictive analytics approach for forecasting bike rental demand. *Decision Analytics Journal*, 11, Article 100482. https://doi.org/10.1016/j.dajour.2024.100482

Kamoona, A., Song, H., Keshavarzian, K., Levy, K., Jalili, M., Wilkinson, R., Yu, X., McGrath, B., & Meegahapola, L. (2023). Machine learning-based energy demand prediction. *Energy Reports*, 9, 171–176. https://doi.org/10.1016/j.egyr.2023.10.071

Kheawpeam, N., & Sinthupinyo, S. (2023). Demand forecasting using machine learning to manage product inventory for multi-channel retailing store. *Proceedings of the IEEE International Conference on Omni-Layer Intelligent Systems (COINS 2023)*. https://doi.org/10.1109/COINS.2023

Klute, B., Homb, A., Chen, W., & Stelpflug, A. (2019). Predicting outpatient appointment demand using machine learning and traditional methods. *Journal of Medical Systems*, 43, Article 288. https://doi.org/10.1007/s10916-019-1418-y

Malik, A., Dargar, G., & Sharma, A. (2023). Predictive analysis for retail shops using machine learning for maximizing revenue. *Proceedings of the 7th International Conference on Intelligent Computing and Control Systems (ICICCS 2023)*. https://doi.org/10.1109/ICICCS56967.2023.10142634

Martins, E., & Galegale, N. V. (2023). Sales forecasting using machine learning algorithms. *Revista Gestão e Secretariado*, 14(7), 11294–11308. https://doi.org/10.7769/gesec.v14i7.1670

Mebal, A. P., Sankar, H., Jothika, S. J., & Manochitra, M. (2021). Predicting the demand for FMCG using machine learning. *International Journal of Engineering and Advanced Technology*, 10(3), 169–174. https://doi.org/10.35940/ijeat.C2253.0210321

Nagaraj, P., Muneeswaran, V., & Dharanidharan, A. (2023). E-commerce customer churn prediction scheme based on customer behaviour using machine learning. *Proceedings of the International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1–10).

Pandey, V. K., Bisoy, S., & Panda, S. (2023). Time series forecasting and prediction of Walmart data using hybrid machine learning techniques. *Proceedings of the International Conference on Circuits, Power, and Intelligent Systems (CCPIS*). https://doi.org/10.1109/CCPIS59145.2023.10291333

Pereira, L. N., & Cerqueira, V. (2022). Forecasting hotel demand for revenue management using machine learning regression methods. *Current Issues in Tourism*, 25(17), 2733–2750. https://doi.org/10.1080/13683500.2021.1999397

Punia, S., & Shankar, S. (2022). Predictive analytics for demand forecasting: A deep learning-based decision support system. *Knowledge-Based Systems*, 258, Article 109956. https://doi.org/10.1016/j.knosys.2022.109956

Rajendran, S., Srinivas, S., & Grimshaw, T. (2021). Predicting demand for air taxi urban aviation services using machine learning algorithms. *Journal of Air Transport Management*, 92, Article 102043. https://doi.org/10.1016/j.jairtraman.2021.102043

Ranjitha, P., & Spandana, M. (2023). Predictive analysis for Big Mart sales using machine learning algorithms. *International Journal of Artificial Intelligence Research*, 7(2), 100–120. https://doi.org/10.1109/ICICCS51441.2023.9432109

Rodrigues, M., Miguéis, V., Freitas, S., & Machado, T. (2024). Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste. *Journal of Cleaner Production*, 435, Article 140265. https://doi.org/10.1016/j.jclepro.2023.140265

Sreekumar, A. V., & Lekshmi, R. R. (2024). Electric vehicle charging station demand prediction model deploying data slotting. *Results in Engineering*, 24, Article 103095. https://doi.org/10.1016/j.rineng.2024.103095

Sousa, M. S., Loureiro, A. L. D., & Miguéis, V. L. (2023). Predicting demand for new products in fashion retailing using censored data. *Expert Systems With Applications*, 259, Article 125313. https://doi.org/10.1016/j.eswa.2024.125313

Sultanbek, M., Adilova, N., Sładkowski, A., & Karibayev, A. (2024). Forecasting the demand for railway freight transportation in Kazakhstan: A case study. *Transportation Research Interdisciplinary Perspectives*, 23, Article 101028. https://doi.org/10.1016/j.trip.2024.101028

Taheri, S., & Razban, A. (2022). A novel probabilistic regression model for electrical peak demand estimate of commercial and manufacturing buildings. *Sustainable Cities and Society*, 77, Article 103544. https://doi.org/10.1016/j.scs.2021.103544

Thakur, A., & Shukla, K. A. (2023). Predictive analysis of energy consumption and electricity demand using machine learning techniques. *Energy Reports*, 9, 567–579. https://doi.org/10.1016/j.energy.2023.01.001

Ulrich, M., Jahnke, H., Langrock, R., Pesch, R., & Senge, R. (2021). Distributional regression for demand forecasting in e-grocery. European *Journal of Operational Research*, 294, 831–842. https://doi.org/10.1016/j.ejor.2019.11.029

Vairagade, N., Logofatu, D., Leon, F., & Muharemi, F. (2019). Demand forecasting using random forest and artificial neural networks. *Proceedings of the International Conference on Computational Collective Intelligence*. Springer.

Van Nguyen, T., Zhou, L., Chong, A. Y. L., Li, B., & Pu, X. (2020). Predicting customer demand for remanufactured products: A data-mining approach. *European Journal of Operational Research*, 281, 543–558. https://doi.org/10.1016/j.ejor.2019.08.015

Vishnu, G., Kaliyaperumal, D., Pati, P. B., Karthick, A., Subbanna, N., & Ghosh, A. (2023). Short-term forecasting of electric vehicle load using time series, machine learning, and deep learning techniques. *World Electric Vehicle Journal*, 14(9), Article 266. https://doi.org/10.3390/wevj14090266

Wang, P., & James, W. (2024). Predicting oil price fluctuations: Integrating external indicators and advanced regression techniques. *Resources Policy*, 97, Article 105263. https://doi.org/10.1016/j.resourpol.2024.105263

Xu, Y., & Li, Q. (2022). Impact of design innovation on consumer preferences in construction materials. *Journal of Civil Engineering and Managemen*t, 28(3), 166–176. https://doi.org/10.3846/jcem.2022.15777

Yan, J., & Liu, J. (2023). Research on Maximum Temperature Prediction Based on ARIMA-LSTM-XGBoost Weighted Combination Model. *Journal of Circuits, Systems, and Computers*. https://doi.org/10.1142/s0218126624501238

Yaneva, P. E., & Kulina, H. N. (2023). Furniture market demand forecasting using machine learning approaches. *Journal of Physics: Conference Series, 2675*, Article 012008. https://doi.org/10.1088/1742-6596/2675/1/012008

Yi, S. (2023). Walmart sales prediction based on machine learning. *Highlights in Science, Engineering and Technology*, 47, 87–90.

Zhang, L., Chen, Y., & Yan, Z. (2023). Predicting the short-term electricity demand based on weather variables using a hybrid CatBoost-PPSO model. *Journal of Building Engineering*, 71, Article 106432. https://doi.org/10.1016/j.jobe.2023.106432